Walter Sievers

Leistungshöhe im Spiegel absoluter Urteile

Experimentelle und theoretische Beiträge zum Problem psychischer Bezugsysteme

1970

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547228_0019

Tag der mündlichen Prüfung: 10.7.1970

Dekan:	Prof.Dr.M.Hellmann
Referent:	Prof.Dr.W.Witte
Korreferent:	Prof.Dr.M.Sader

UNIV.
LEIDEN
BIBL.

Phototechnische Zentralstelle
der Westfälischen Wilhelms-Universität
zu Münster/Westf.

Inhaltsverzeichnis.

1. Gütemaßstäbe und Bezugsysteme. 4

2. Allgemeines über Bezugsysteme. 10
   (a) Formale Aspekte. 10
   (b) Beliebigkeit bzw. Nicht-Beliebigkeit von Eigenschaften-Systemen. 12
   (c) Systemkonstituierende Ereignisräume. 14

3. WITTEs Theorie der Bezugsysteme. 20
   (a) Allgemeine Annahmen. 20
   (b) Spezielle Annahmen. 23

4. Einige Überlegungen zur Verteilungsform kategorialer Urteile. 26

5. Auswahl des Gegenstands-Systems und des experimentellen Vorgehens. 30
   (a) Spezifität der Trägermenge. 30
   (b) Experimentelles Vorgehen. 33
   (c) Auswahl der Merkmalskontinua. Die Untersuchungen von KELLNER. 35

6. Gewinnung einer Skala. 52
   (a) Allgemeiner Begriff einer Skala. 52
   (b) Abbildungsverhältnisse bei einer Kategorien-Skala. 53
   (c) Metrische Axiomen-Systeme. 58
      i. PFANZAGLs Axiomen-System. 58
      ii. Infinites - finites Axiomen-System (SUPPES & ZINNES). 63
   (d) Skalierung durch abgeleitete Messung und Berechnung der Systemwerte. 65

7. Befunddiskussion. 73
   (a) Zufallskritische Überprüfung der Differenzen zwischen empirischen und theoretischen Systemwerten. 74
   (b) Normalverteilung kategorialer Urteile. 78

(c) Zum Begriff des "mnemisch stabilisierten Systems". 80

(d) Verteilungsform kategorialer Urteile. 87

(e) Invarianz von Distanzen bei variierendem Verteilungstyp (PARDUCCI). 88

(f) Invarianz von Distanzen bei variierender Position der Beurteiler. 92

(g) Mitgliedsgruppe - Bezugsgruppe. 102

8. Beurteilungen über ein nicht-anschauliches Leistungsmerkmal. 107

9. Vergleich Kategorien-Skalen - Rangreihen. 110

Anhang. 115

Literaturverzeichnis. 126

1. Gütemaßstäbe und Bezugsysteme.

In seinem Beitrag zum Tübinger Symposion über Motivation: "Eine Rahmentheorie der Motivation in zehn Thesen" spricht HECKHAUSEN (11) von Motivationen als beruhend "auf hochorganisierten kognitiven Systemen von Person-Umwelt-Bezügen" (These 1). "Jedes System stellt mit seinem verallgemeinerten Grundgerüst ein Bezugsystem für den jeweilig aktuellen Person-Umwelt-Bezug dar. Es legt insbesondere fest, wie ein Umwelt-Bezug beschaffen sein muß, um sich als ein befriedigender vor anderen auszuzeichnen" (These 4).

Das soll heißen: "Da es sich bei den Motivationsphänomenen im allgemeinen Sinne um Regulationen des Person-Umwelt-Bezuges von seiten der Person handelt, entsteht zunächst die Frage, woraufhin denn reguliert wird. Das wird hier vage als 'Norm' angesprochen, die festlegt, was als ein 'befriedigender' Bezug ... zu gelten hat. So kann die leistungsthematische Norm eines Menschen darin bestehen, der Spitzen- oder Mittelgruppe der Leistungs-Rangverteilung einer relevanten Bezugsgruppe anzugehören. Die Norm bestünde dann in einem sozialbezogenen hochgeneralisierten Anspruchsniveau von bestimmter Höhe ..." (Kommentar zu These 4).

HECKHAUSEN gibt dann weiterhin (These 5) fünf "Faktoren" an, durch die die "Norm eines befriedigenden Person-Umwelt-Bezuges ... festgelegt werden kann", von denen wir die für den späteren Verlauf der Argumentation drei möglichen relevanten herausgreifen:
1. Der "Niederschlag der diesbezüglich bislang erfah-

renen Lebenslagen";
2. die "sachimmanente Prägnanzstruktur einer Inhaltsklasse von Person-Umwelt-Bezügen";
3. die "Übernahme tradierter Norm-Vorstellungen".

Weiterhin: "Je nach dem Gewicht dieser Faktoren können Soll-Lagen verschiedener Natur sein: etwa durchschnittsstatistische Normal-Lagen nach Art sich wandelnder Adaptationsniveaus (wenn Erfahrungsniederschläge ausschlaggebend sind); ... ; oder Ideal-Lagen, die weniger ein fixiertes Niveau als eine Richtungstendenz, ein Gefälle auf einen ausgezeichneten Prägnanzzustand hin darstellen (wenn sachimmanente Prägnanzstrukturen oder tradierte Normvorstellungen ausschlaggebend sind) ..." (These 6).

Wenn HECKHAUSEN im Kommentar zu These 4 von "der Spitzen- oder Mittelgruppe der Leistungs-Rangverteilung einer relevanten Bezugsgruppe" spricht, der eine Person angehört bzw. angehören möchte, so setzt das voraus, daß Leistungen in eine "Rangverteilung" gebracht werden können derartig, daß über eine bloße Anreihung einer Rangverteilung hinaus sich auch noch Spitzen-, Mittel- und andere Gruppen herausgliedern können.

An anderer Stelle sagt HECKHAUSEN (10, S. 604): "Leistungsmotivation läßt sich ... definieren als das Bestreben, die eigene Tüchtigkeit in allen jenen Tätigkeiten zu steigern oder möglichst hoch zu halten, in denen man einen Gütemaßstab für verbindlich hält und deren Ausführung deshalb gelingen oder mißlingen kann." - "Relevante Ereignisse eines leistungsthematischen Person-Umwelt-Bezuges - und damit zeitliche Endzustände, auf die sich Leistungsmotivation bezieht - sind letzlich immer Erfolg und Mißerfolg.

Wie differenziert auch ein Gütemaßstab sein mag, er zerfällt in einen Teil, der -wie ausgeprägt auch immer- Erfolg, und in einen anderen, der Mißerfolg bedeutet ... In Abhängigkeit vom erzielten Leistungsstand wandert ... der konkretisierte Nullpunkt des Gütemaßstabes."

Anstelle der, in der Klassifikation der Skalen, "schwachen" Skala einer Leistungsrangverteilung wird nun hier von einem Gütemaßstab gesprochen, der verschiedene Grade der Differenziertheit, d.h. der Meßgenauigkeit, haben kann und seinen Ursprung in einem Nullpunkt hat. Einem solchen Gütemaßstab werden also Eigenschaften zugesprochen wie üblicherweise einer "Verhältnis"-Skala (oder, in der Terminologie PFANZAGLs (22), einer Skala, die eindeutig ist bis auf Streckungen).

Vereinfachend gesprochen läßt sich das von HECKHAUSEN Gemeinte so darstellen: Ergibt sich innerhalb eines "leistungsthematischen Person-Umwelt-Bezuges" ein Leistungsspektrum (etwa auf dem Gebiet sportlicher Leistungen), so wird mit Hilfe eines Maßstabes die Güte der auftretenden Leistungen gemessen, sodaß die jeweilige "messende" Person den Ort ihrer eigenen Leistungen auf diesem Maßstab und den Abstand zu Leistungen anderer feststellen kann. Je nach der Gesamtkonstellation ihrer Leistungsmotivation ist sie dann mit diesem Ort zufrieden, oder sie bemüht sich, den Abstand zu anderen Leistungen in der gewünschten Richtung hin zu verringern.

Diese Leistungen müssen nun nicht als solche gesehen werden, sondern als durch jeweilige Personen repräsentiert, wenn es Sinn haben soll, von einer Zugehörigkeit zu einer "Spitzen- oder Mittelgruppe einer Leistungs-Rangverteilung" zu sprechen.

Wenn aber die Zuordnung von Personen bezüglich einer thematischen Leistung nur innerhalb einer Bezugsgruppe zur Spitze oder zur Mitte möglich ist, dann ist der Zusammenhang zwischen der Frage nach Gütemaßstäben und der nach Bezugsystemen unmittelbar gegeben.

Auf diesen Zusammenhang weist HECKHAUSEN in seinem Symposion-Beitrag auch ausdrücklich hin nach einer Klassifikation der Soll-Lagen in These 6: "Bei der Klassifikation der Arten von Soll-Lagen geht es um das allgemeinpsychologische Problem der Bezugsysteme und ihrer Genese, das für die Motivationspsychologie nicht weniger grundlegend ist als für die Wahrnehmungspsychologie ... Die Genese von Gütemaßstäben mit ihren Gradabstufungen und Soll-Lagen ist bislang kaum erforscht. Vielleicht läßt sich der genius loci (Tübinger Symposion, 1962 (!)) der WITTEschen Bezugsystem-Forschung beschwören, sich dieser motivationspsychologischen Fragen anzunehmen." (+)

Die dadurch gestellte Aufgabe ist nun freilich von einer außerordentlichen Komplexität nicht nur hinsichtlich der Schwierigkeit des Problems durch das Ineinandergreifen sehr vieler Komponenten und auch oft hinsichtlich der Gewinnung relevanter Daten, sondern auch durch die Mannigfaltigkeit der auftretenden Fragestellungen mit allen ihren Verästelungen.

Ist die Rede von leistungsmotivationalen Gütemaßstäben, so lassen sich hierbei zwei Aspekte unterscheiden:

(+) Der an dieser Stelle von HECKHAUSEN gestiftete Zusammenhang von Gütemaßstäben und WITTEscher Bezugsystem-Forschung lieferte die Anregung für diese Arbeit.

I. Der eher leistungsmotivationale Aspekt mit den Fragen etwa nach Genese von Gütemaßstäben (diese Frage ist nicht unbedingt identisch mit einer möglichen entwicklungspsychologischen), nach möglichen Wechselwirkungen zwischen Motivationshöhe und Differenziertheit des Maßstabes, seinen Veränderungen bei Veränderungen des Leistungsniveaus u.a. Also mehr "inhaltliche" Fragestellungen, die man durch das Titelschema: der Einfluß von ... auf ... charakterisieren könnte.
II. Der eher "formale" Aspekt des Bezugsystems mit den Fragen etwa nach System-"Umfang" (was das auch immer heißen möge), nach Metrik, Gliederung von Systemen, Verhältnis verschieden differenzierter Systeme zueinander (wie auch immer) u.a.

Der zweite Aspekt hat logisch einen Vorrang insofern, als es wenig sinnvol erscheint, über dynamische Aspekte von Gütemaßstäben zu reden, bevor nichts über ihre Metrik bekannt ist. Die Frage der Metrik ist aber (wie oben gezeigt) nicht unabhängig diskutierbar von der Frage nach den zugrunde liegenden Bezugsystemen und den damit zusammenhängenden Fragen.

Damit ist der Weg möglicher Untersuchungen zum Thema Gütemaßstäbe gegeben. Wir werden uns in dieser Arbeit im wesentlichen auf die "formalen" Aspekte von Bezugsystemen beschränken und nicht mehr auf den "inhaltlichen" Kontext zur Leistungsmotivation zurückkommen.

Diese Einschränkung hat ihren Grund allerdings auch darin, daß von der WITTEschen Bezugsystem-Forschung her, unabhängig von den Fragestellungen

der Leistungsmotivation (bzw. allgemein der Motivation), eine autochthone Fragestellung besteht. In den bisherigen Untersuchungen zur Theorie der Bezugsysteme wurden durchweg (von einer Ausnahme abgesehen, auf die wir später zurückkommen als der unmittelbaren Vorgängerin dieser Arbeit) die Systembildungen an Eigenschaften von Gegenstandsmannigfaltigkeiten untersucht, die über eine physikalische Metrik (cm, gr, Hz z.B.) anordbar waren. Die durch das psychologische Experiment gefundenen phänomenalen Relationen wurden dann über dieser Metrik abgetragen+. Die Abbildung dieser phänomenalen Relationen auf die Metrik der vorgeordneten Gegenstände wurde als wenigstens monoton, meist als linear angenommen. Vergleiche verschiedener Systeme über dieselbe Eigenschaft konnten über diese Metrik vorgenommen werden.

Die Frage war nun, ob die bisher gefundenen Verhältnisse (auf die Details dieser "Verhältnisse" wird später noch eingegangen) sich reproduzieren lassen, wenn man die Frage nach Bezugsystemen erweitert auf Eigenschaften oder Gegenstandsbereiche, innerhalb derer eine physikalische Metrik nicht angebbar ist. (Vielleicht sollte man vorsichtiger sagen, daß ein "einfaches" physikalisches Maß nicht angebbar ist, oder, wenn schon nicht einfach, so doch ein bisher nicht bekanntes Maß.) Dieser Schritt entspricht dem in der Geschichte der Psychologie, als nach der Periode der älteren, klassischen Psychophysik die Frage nach der Möglichkeit der Skalierung "psychologischer" Gegenstände (z.B. Einstellungen) auftauchte.

+ Streng genommen bei Tonhöhen nicht.

2. Allgemeines über Bezugsysteme.

(a) Formale Aspekte.

Redet man von einem Bezugsystem, so ist damit angesprochen eine Zuordnung von etwas zu etwas als etwas derartig, daß diese Zuordnung als etwas nur möglich ist relativ auf eine andere Zuordnung. Unter Berücksichtigung dieses Sachverhalts läßt sich folgende Menge Mg angeben:

$$Mg\ \{\{a,b\},\ \{r,s\}\}$$

derartig, daß

$$a \longrightarrow r,\ b \longrightarrow s$$

zugeordnet wird.
Ist in der Menge Mg $\{a,b\}$ eine Relation R angebbar, sodaß aRb, dann

$$(a \longrightarrow r \wedge b \longrightarrow s) \wedge aRb \not\Longrightarrow rRs\ .$$

D.h. ist eine Relation R gegeben, sodaß aRb, dann muß die Zuordnungsmenge r,s in ihrer Zuordnung zu Mg $\{a,b\}$ als etwas diese Relation schon darstellen. Dieses verlangt aber nicht, daß bei verbaler Darstellung der Zuordnungsmenge diese eine "komparative" Form annimmt, da die Zuordnung als "als etwas" angegeben ist. Die beiden Elemente der Zuordnungsmenge (also die Eigenschaften) sind polar angeordnet. Die Minimalbedingung eines Bezugsystems läßt sich also darstellen als die Menge zweier geordneter Paare

$$Mg\ \{\langle a,r\rangle,\ \langle b,s\rangle\}$$

wobei $a,b \in T$, Trägermenge einer Eigenschaft
$r,s \in E$, Menge der polar geordneten Eigenschaften.

Die Zeichen bedeuten: $\not\Longrightarrow$ für logische Implikation
$\wedge$ für logischen Produkt("und")
$\in$ Elementbeziehung

Hierzu ein Beispiel:
Es seien zwei Kreise a,b gegeben (die Menge aller Kreise sei zweielementig), eine Relation R = ... ist größer als ..., so daß aRb, und eine Zuordnungsmenge Mg(Z) = { r,s } derartig (da es sich hier um Zuordnungen von etwas als etwas handelt, also Eigenschaften), daß ihre Elemente in einem Zuordnungskontext so zu bestimmen sind, daß das eine dem größeren, das andere dem kleineren Element der Menge der Kreise zugeordnet wird, so daß bei der angegebenen Relation R r = "groß", s = "klein" bedeuten kann. Die Zuordnungen a ⟶ r, b ⟶ s heißen dann "a ist groß", "b ist klein".

Dieses konstruierte Beispiel ist freilich rein formal gedacht. Im Alltag würde wohl tatsächlich unter derartigen Bedingungen der eine Kreis als "der Große", der andere als "der Kleine" bezeichnet werden, wie es nicht selten z.B. in Familien mit zwei Jungen passiert, wo der ältere eben dann "der Große" ist, der jüngere "der Kleine", und beide unter Umständen diese Bezeichnung trotz aller Größenwandlungen lange Zeit ihres Lebens beibehalten.

Zu einem Bezugsystem gehören also eine Menge T von Gegenständen, die in irgendeiner Weise zusammengefaßt sind, die Trägermenge einer Eigenschaft, und eine Relation R in dieser Menge, über die sie geordnet werden kann. Das geordnete Paar $\mathcal{T} = \langle T,R \rangle$ wäre dann ein Gegenstands-System. Weiterhin gehört dazu ein Eigenschaften-System $\mathcal{E} = \langle E,R' \rangle$, wobei E gleich Menge der Eigenschaften (Zuordnungsmenge), die als polar angeordnet gedacht ist über die Relation R'. Außerdem die Zuordnung Z. Ein Bezugsystem ist also ein geordnetes Tripel $\mathcal{B} = \langle \mathcal{T}, \mathcal{E}, Z \rangle$.

Diese letztere Darstellung in Anlehnung an die Definition einer Skala bei SUPPES & ZINNES(28). Es ist allerdings zu beachten, daß wir hier noch keine Skala definiert haben.

Die Charakterisierung des Eigenschaften-Systems ist etwas schwierig, da die Eigenschaften für sich genommen ja nicht in einer Relation stehen. Die Eigenschaft "groß" ist nicht "größer" als die Eigenschaft "klein", sondern über ihre allgemeine semantische Funktion sind "groß", "klein" ordbar dadurch, daß innerhalb eines Systems durch "groß" eben größere Gegenstände bezeichnet werden als durch "klein". $\mathcal{T}$, $\mathcal{E}$, Z sind als Komponenten eines Systems nicht unabhängig voneinander definierbar.

Den Begriff "Eigenschaft" haben wir im übrigen bisher zweifach verwendet. Einmal im Sinne einer Merkmalsdimension neben anderen eines Gegenstands, zum anderen im Sinne eines Ausprägungsgrades dieser Merkmalsdimension. Im weiteren Verlauf soll für letzteren Fall der Begriff der Eigenschaft vorbehalten bleiben. Bei der besprochenen Menge der Eigenschaften E des Eigenschaften-System $\mathcal{E}$ handelt es sich um Eigenschaften im Sinne dieses letzteren Falles.

(b) Beliebigkeit bzw. Nicht-Beliebigkeit von Eigenschaften-Systemen.

Es ließe sich die Klasse Kl aller polar geordneten, vollständigen Eigenschaften-Systeme Kl $\{\mathcal{E}_i\}$ konstruieren und zu jedem $\mathcal{E}_i$ ein Körper $\mathfrak{Q}_i$ mit allen möglichen Teilmengen $\mathcal{E}_{it}$ von $\mathcal{E}_i$ als Elementen, die polar und symmetrisch geordnet sind. Haben wir ein Bezugsystem $\mathcal{B}_i = \langle \mathcal{T}_i, \mathcal{E}_i, Z_i \rangle$, dann ist jedes $\mathcal{E}_{it}$ aus $\mathfrak{Q}_i$ eine Zerlegung von $\mathcal{T}_i$. Je umfangreicher die Teilmengen $\mathcal{E}_{it}$, desto feiner die Zerlegung von $\mathcal{T}_i$.

Auf diese Weise ließe sich die Behauptung von der "Relativität der Urteile" und die verbreitete experimentelle Praxis, alle möglichen Eigenschaften-Systeme $\mathcal{E}_{it}$ zur Beurteilung gegebener Trägermengen T als gleichmöglich zu betrachten (wie etwa bei HELSON und vielen anderen, die auf denselben oder ähnlichen Gebieten arbeiten wie er), -formalisieren. Wir wollen diese Behauptung die von der "Beliebigkeit des (natürlich konformen) Eigenschaften-Systems $\mathcal{E}_{it}$ " nennen.

Betrachten wir ein bei Kategorisierungsexperimenten weit verbreitetes Eigenschaften-System von Größen-Eigenschaften:⟨sehr sehr groß, sehr groß, groß, eher groß, mittel, eher klein, klein, sehr klein, sehr sehr klein⟩. Die Eigenschaften sind offensichtlich polar und symmetrisch geordnet. Die o.a. Bedingung der Vollständigkeit ist allerdings nicht leicht zu erfüllen, da sich durch Kompositionen immer extremere Eigenschaften bilden lassen einerseits, andererseits die Eigenschaften immer weniger "trennscharf" werden, d.h. in ihrer semantischen Funktion ungenügend gegeneinander abgrenzbar sind, und damit zusammenhängend ihre Ordnung nicht unmittelbar ablesbar ist. Aus dem allgemeinen Wortschatz ließe sich aber die Serie an dem einen Extrem erweitern durch Eigenschaften wie: außerordentlich groß, riesig, gewaltig, ..., unermeßlich (bezeichnenderweise!) groß usw.; zum anderen Extrem durch: außerordentlich klein, winzig, verschwindend klein usw. Polar geordnete, symmetrische Teilmengen eines solchen Eigenschaften-Systems wären dann z.B. folgende Eigenschaften-Systeme: ⟨groß, klein⟩, ⟨sehr groß, groß, mittel, klein, sehr klein⟩.

(c) Systemkonstituierende Ereignisräume.

Die Tatsache, daß es im allgemeinen weder winzige Gebirge, noch gewaltige Sandkörner gibt und anscheinend nicht einmal sehr große Bleistifte (s. WITTE (36)), sich andererseits aber Gegenstände aufweisen lassen, denen derartige Eigenschaften nachgesagt werden können (etwa die "gewaltige" Entfernung zwischen den Sternen, "winzige" Mikroben, "sehr große"Bäume), läßt es unvermeidlich erscheinen, einige Annahmen zu machen, die zunächst recht spekulativ erscheinen mögen, aber doch einigen Erklärungswert besitzen. Der ganze hier nun angesprochene Problembereich ist freilich von einer außerordentlichen Komklexität, sodaß es jetzt nur möglich ist, einige Teile aus ihm zu skizzieren.

Läßt sich selbst von den größten üblicherweise verwendeten Bleistiften nicht sagen, daß sie "sehr groß" sind, sondern allenfalls "groß", so weist das offensichtlich darauf hin, daß das System "Bleistift" (bezüglich der Länge dieser) nicht in sich abgeschlossen ist, sondern in irgendeiner Weise selbst wieder zu sehen ist auf der Folie der räumlichen Extensitäten aller anderen Gegenstände. Das System Bleistift wäre dann Teilsystem eines umfassenderen Systems aller möglichen Längen.

Diese Erklärung reicht aber nicht aus für die Tatsache, daß es keine "sehr großen" Bleistifte gibt. Denn im Verhältnis zur Größe aller möglichen anderen Gegenstände und Abstände liegt der Sektor des Systems Bleistift auf dem Kontinuum aller möglichen in cm meßbaren Längen am

unteren Ende, sodaß es von daher sogar verwunderlich erscheinen könnte, daß es überhaupt "große" Bleistifte gibt. Offensichtlich sind sie aber nicht nur Teilsystem des Systems aller möglichen Gegenstände und Abstände, das durch "Langheit" konstituiert wird, sondern, irgendwie vermutlich immer, Teilsystem von Systemen anderer systemkonstituierender Eigenschaften (Merkmale). Durch jeweils gegenstandsspezifische Merkmalskomplexionen kann sich dann eine betrachtete Menge von Gegenständen bezüglich eines betrachteten Merkmals relativ herauslösen aus dem umfassenden Merkmalskontinuum, sodaß einzelne Elemente dann Eigenschaften annehmen können, die ihnen im Rahmen des umfassenden Systems nicht zuständen. Zum anderen besteht durch die Tatsache, daß jeder Gegenstand durch einen Merkmalskomplex beschrieben werden kann, die Möglichkeit, daß über eine Teilmenge von Merkmalen der Menge dieses Komplexes Systemmengen konstituiert werden, bezüglich derer dann ein Gegenstand hinsichtlich des betrachteten Merkmalskontinuums originär eine Eigenschaft annehmen kann. Das umfassende System wäre dann nicht primär die Menge aller möglichen Gegenstände, die Trägermenge eines betrachteten Merkmals ist, auf dessen Kontinuum die betrachtete Gegenstandsmenge einen Sektor darstellt, sondern die Menge der Gegenstände, die als "Verwandte" anzusprechen sind.

Zum Beispiel: als derartige Verwandte von Bleistiften könnte man alles "Schreibzeug" betrachten wie Kugelschreiber, Radiergummi, Federhalter, Lineal, L$_o$cher, Hefter, Schreibpapier usw.,also alle möglichen Dinge, die sich üblicherweise auf einem Schreibtisch befinden. Gleichzeitig ist es möglich, daß über ein anderes Merkmal (sc. als Schreibzeug) eine "Familie" von Gegen-

ständen erfaßt wird, die systemkonstituierenden Charakter besitzt. Auf diese Weise erhält man ein "System" von Systemverschachtelungen, innerhalb dessen eine Hierarchie der Systeme bestehen mag und/oder eine Nebenordnung hinsichtlich der Stärke der Wirksamkeit bei der Ausbildung einer Eigenschaftenstruktur eines Merkmalskontinuums einer betrachteten Menge von Gegenständen. Diese Wirkungen können sich dann addieren oder neutralisieren.

Das umfassendste System hätte als Trägermenge das Universum aller möglichen Dinge. Dieses Universum der Dinge kann man auffassen als dargestellt durch eine Produktmenge AxBxCx ... xN des Universums aller möglichen Merkmale A,B,C, ... ,N . Für jedes Merkmal ließe sich ein Ereignisraum angeben, dessen Ereignisse die Menge aller möglichen Ausprägungsgrade des Merkmals seien. Dann stellt die Produktmenge über das Universum der Merkmale den universalen Ereignisraum dar. Jedes Element der universalen Trägermenge läßt sich somit definieren durch ein geordnetes N-tupel (a,b,c, ... ,n / a,b,c, ... ,n sind Ausprägungsgrade der Merkmale A,B,C, ... ,N). Fassen wir A,B,C, ... ,N als Zufallsveränderliche auf, die als Definitionsbereich die Menge ihrer Ausprägungsgrade haben, dann stellt das System der Merkmale einen zufälligen Vektor dar, dessen Verteilungsfunktionen durch die Trägermenge gegeben ist. Ebenso lassen sich dann natürlich für die einzelnen Merkmale Verteilungsfunktionen angeben.

Die beiden letzteren Behauptungen idealisieren freilich die realiter vorliegenden Verhältnisse. Genaugenommen lassen sich die Verteilungsfunktionen natürlich immer nur für einen gegebenen Zeitpunkt t_i angeben, da die Trägermenge im ständigen Wandel begriffen ist.

Diese universale Verteilungsfunktion eines universalen Ereignisraumes ist freilich eine nur denkbare Konstruktion. Tatsächlich gegeben und relevant für eine Diskussion im Rahmen der Psychologie ist aber nur derjenige Ereignisraum, der durch die jeweilige individuelle Erfahrung konstituiert worden ist. Wir werden entsprechend diesen Ereignisraum im Unterschied zum universellen den (individuellen) Erfahrungsraum nennen, für den dann wieder eine individuelle Verteilungsfunktion besteht. Bezugssysteme bilden sich über diese Erfahrungsräume aus.

Die Konstruktion des universellen Ereignisraumes hat ihre Berechtigung dadurch, daß es möglich wird (a) interindividuelle Differenzen von Bezugsystemen aufzuklären, die bedingt sind durch differierende Erfahrungsräume und zugehörige Verteilungsfunktionen, die dann als Stichproben des universellen Ereignisraumes und seiner zugehörigen Verteilungsfunktionen aufgefaßt werden können; (b) Verschiebungen von Strukturen ausgebildter Bezugsysteme durch das Auftreten von Ereignismöglichkeiten des universellen Ereignisraumes, die individuell außerhalb des bisherigen Ereignisraumes lagen (d.h. wenn Gegenstände bekannt werden, die hinsichtlich des Ausprägungsgrades eines Merkmals bisher unbekannt waren), oder/und durch Veränderungen der Verteilungen über den bisherigen Erfahrungsraum.

Bei relativ stabilen Umweltbedingungen können die individuellen Verteilungsfunktionen derjenigen Individuen, die innerhalb einer relativ gemeinsamen Umwelt leben, bei Konvergenz gegen die universelle Verteilungsfunktion gegeneinander konver-

gieren. Dadurch besteht die Möglichkeit interindividuell gemeinsamer Bezugsysteme, d.h. dann interindividueller Übereinstimmung in der Zuordnung von Elementen der Trägermenge zu Eigenschaften. Das Maß dieser Übereinstimmung kann allerdings nur experimentell festgestellt werden.

Es scheint sinnvoll,noch einen weiteren Ereignisraum anzunehmen und zwar den durch das Experiment vorgegebenen Ereignisraum mit der zugehörigen Verteilungsfunktion. Nennen wir diesen zur Unterscheidung von den anderen den Experimentalraum.

Wir haben also nun drei Ereignisräume unterschieden:

1. Den universellen Ereignisraum,
2. die (individuellen) Erfahrungsräume, die gegebenenfalls gegeneinander konvergieren können,
3. den Experimentalraum.

Der Erfahrungsraum und der experimentell vorgegebene Ereignisraum können von vornherein identisch sein für den Fall einer konstruierten Trägermenge, etwa Striche, Quadrate, Kugeln.

Bei experimentellen Untersuchungen über Bezugsysteme kann zunächst nur der Ereignisraum angenommen werden, den man experimentell vorgegeben hat. Untersucht man Systeme, die sich über Trägermengen ausgebildet haben, die aus dem Alltag bekannt sind, und liegt der experimentell vorgegebene Ereignisraum nur auf einem Sektor des einschlägigen Merkmalskontinuums, so wird man keine Abbildung auf das gewählte polar geordnete Eigenschaften-System erreichen (vorausgesetzt, daß man nicht

Null-Abbildungen als zulässige Abbildungen im Geltungsbereich eines Bezugsystems annehmen will). Um ein triviales Beispiel anzugeben: werden zur Untersuchung des Systems "Bleistift" nur Bleistifte ausgewählt, die längenmäßig im oberen Drittel der tatsächlich existierenden Bleistifte liegen, so wird es kaum möglich sein, "klein-" bzw. "kurz-" Urteile zu erhalten, da sich das System über ein größeres Kontinuum von Längen ausgebildet hat.

HELSONs Formel für das Adaptations-Niveau (AN):

$$AN = X^p X_r^q X_r^r \quad ,$$

wobei $$p + q + r = 1 \quad ,$$

welche das AN als gewichtetes geometrisches Mittel von Serienreizen (X), die gerade zur Beurteilung vorliegen, sowie von Hintergrundsreizen (X_h) und Residualreizen (X_r) interpretiert, reflektiert den Sachverhalt verschiedener Ereignisräume.

> Freilich interpretiert HELSON sie nicht als solche. HELSON nimmt außerdem von vornherein an, daß die verschiedenen Ereignisräume einen "Kompromiß" eingehen, wenn auch die 2. und/oder 3. Komponente gleich Null gesetzt werden kann - über eine log-Darstellung der Gleichung -, was berechtigt ist unter der Bedingung psychophysischer Experimente mit "sinnfreien" Gegenständen.

Die Annahme verschiedener Ereignisräume, speziell des Erfahrungsraumes und des experimentell vorgegebenen Ereignisraumes, scheint aber exakter und Basis für größere Verallgemeinerungen und Spezifikationen im einzelnen Fall zu sein, auch hinsichtlich der Möglichkeit mehrdimensionaler Systeme. Sie gestattet auch, eine Interpretationsgrundlage zu schaffen für Versuchsergebnisse, die z.B. bei HELSON mit so gewagten Begriffen wie dem des "mental anchor" erklärt werden.

3. WITTEs Theorie der Bezugsysteme.

(a) Allgemeine Annahmen.

Bei Bezugsystemen war bisher gesprochen worden von Zuordnungen von über eine thematisierte Eigenschaft (die systemkonstituierende Eigenschaft, Merkmal) geordnete Menge von Gegenständen zu einer geordneten Menge von Eigenschaften (den Systemeigenschaften). Diese Zuordnung (bzw. "Abbildung") sei monoton. Es läßt sich nun fragen, wie diese Zuordnung geschieht.

Ist die betrachtete Trägermenge geordnet, so hat sie (unter der Annahme, daß sie endlich ist - den Fall unendlicher Mengen braucht man in der Psychologie im allgemeinen wohl nicht zu diskutieren -) ein erstes und ein letztes Element. Findet die Anordnung der Elemente der Trägermenge durch Ähnlichkeitsrelation bezüglich des thematisierten (systemkonstituierenden) Merkmals statt, so besteht das spezifisch Bezugsystemhafte darin, daß das erste und das letzte Element der Trägermenge die Bezugspunkte sind, auf die hin sich die Ähnlichkeitsrelationen ordnen. Die Ähnlichkeit zum ersten Element nimmt vom ersten zum letzten Element hin ab und umgekehrt. Die Ähnlichkeitsabnahme (bzw. -zunahme) ist linear zu der in dieser Weise geordneten Mannigfaltigkeit der Elemente. Für die geometrische Konstruktion dieses linearen Ähnlichkeitsmodells s. Abb.1. Die Ähnlichkeitsdistanz vom ersten zum letzten Element der geordneten Trägermenge läßt sich auf das Intervall $[0,1]$ normieren.

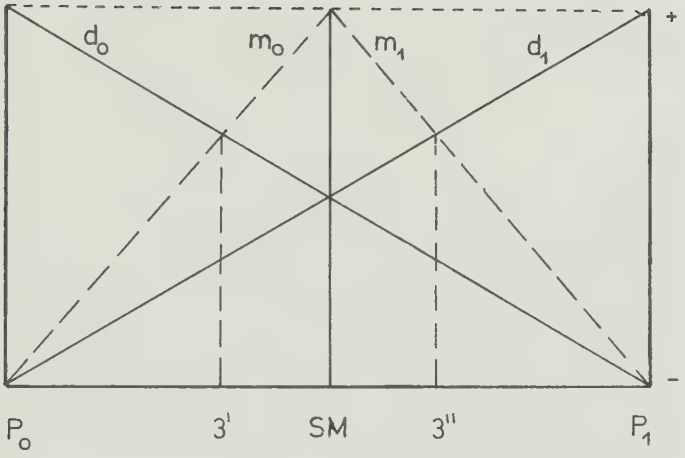


Abb. 1. Geometrische Konstruktion des System-Modells nach WITTE.

d_o, d_1 Ähnlichkeitsgradienten zu den Polen P_o, P_1
m_o, m_1 Mitten-Ahnlichkeiten bezüglich P_o, P_1
SM System-Mitte
3', 3'' System-Drittel (1/3, 2/3)
+, - maximaler, minimaler Ähnlichkeitsgrad

Die Gerade $\overline{P_o P_1}$ ist zu verstehen als das betrachtete Ähnlichkeitskontinuum, auf dem sich die Elemente einer gegebenen Trägermenge anordnen.

Bilden die beiden Randelemente einer gegebenen Trägermenge die Bezugspunkte für die Ähnlichkeitsrelationen in dieser Menge, so kann man sie die Pole (P_o, P_1) des Systems nennen. Die Zuordnung (bzw. Abbildung) der Elemente der Trägermenge auf die Elemente der polar und symmetrisch geordneten Zuordnungsmenge (bzw. Bildmenge) geschieht nun so, daß bei Zuordnung der beiden Polelemente der Trägermenge zu den Polelementen der Zuordnungsmenge (bzw. hat die Zuordnungsmenge nur zwei Elemente, dann auf je eins der beiden Elemente) die übrigen Elemente der geordneten Trägermenge zu den Elementen der Zuordnungsmenge über die überwiegende Ähnlichkeit zu den Polelementen zugeordnet werden. Die Elemente der Zuordnungsmenge (also die (System-) Eigenschaften)können jeweils nur interpretiert werden über dieses Zuordnungsverhältnis zu einer Trägermenge, die derartig ist, daß zwei Elemente aus ihr Poleigenschaften annehmen und die übrigen Elemente sich über Ähnlichkeit zu diesen Polen anordnen lassen. Dieses ist in wesentlichen Zügen das <u>funktionale</u> Ähnlichkeitsmodell für Bezugsysteme nach WITTE in den allgemeinen Annahmen.

Man könnte sich nun überlegen, ob es nicht möglich wäre, daß innerhalb einer Trägermenge mehrere Elemente Poleigenschaften annehmen, sodaß die Elemente dieser Menge sich nach verschiedenen Ähnlichkeitsrelationen R_i (i = 1,2,3, ... ,n) anordnen lassen. Die Menge der Paare von Polen $(P_{i0},P_{i1})_{(i=1,2,3,...,n)}$ spannen dann einen n-dimensionalen "Bezugskörper" auf. Das WITTEsche Modell wäre dann als eindimensional anzusprechen. Es scheint aber so zu sein, daß Bezugsysteme sich immer nur über <u>ein</u> Merkmal ausbilden, sodaß die Elemente der Trägermenge über dieses Merkmal linear angeordnet werden können. Dieses schließt aber nicht aus, daß ein gegebenes Merkmal mannigfache Aspekte besitzt und als Zusammenfassung dieser Aspekte

aufgefaßt werden kann, sodaß sich der Ort eines Elements der Trägermenge auf gedachter Merkmalsgeraden sich ergibt etwa als gewogenes Mittel der merkmalskonstituierenden Aspekte.

(b) Spezielle Annahmen.

Geschieht die Zuordnung der Elemente der Trägermenge zu den Elementen der Bildmenge nach der Ähnlichkeit zu den Polelementen, so heißt das, daß bei angenommenem linearen Ähnlichkeitsgefälle von Pol zu Pol die jeweils größere Polähnlichkeit über die Zuordnung entscheidet und das Intervall [0,1] durch die Zuordnung im Falle einer zweielementigen Bildmenge in zwei gleich große (halboffene) Intervalle [0,1/2), (1/2,1] geteilt wird. Die entsprechende Mitte kann nun selbst wieder Organisationspol einer Mittenähnlichkeit werden, welche nach beiden Randpolen wieder linear abnimmt, sodaß bei Zuordnungen zu einer dreielementigen Bildmenge das Intervall [0,1] in drei gleich große Intervalle unterteilt wird. Die Konstruktion läßt sich weiterführen für noch mächtigere Bildmengen. Abb. 2 veranschaulicht die Konstruktionen für Bildmengen mit 2,3,4,5 Elementen samt einer experimentellen Realisierung des Modells (über die Details des Experiments s. Anhang I). Die durch die Zerlegung der Trägermenge durch die Bildmenge entstehenden Intervalle als Teilintervalle des normierten Intervalls [0,1] werden "Bereiche" genannt.

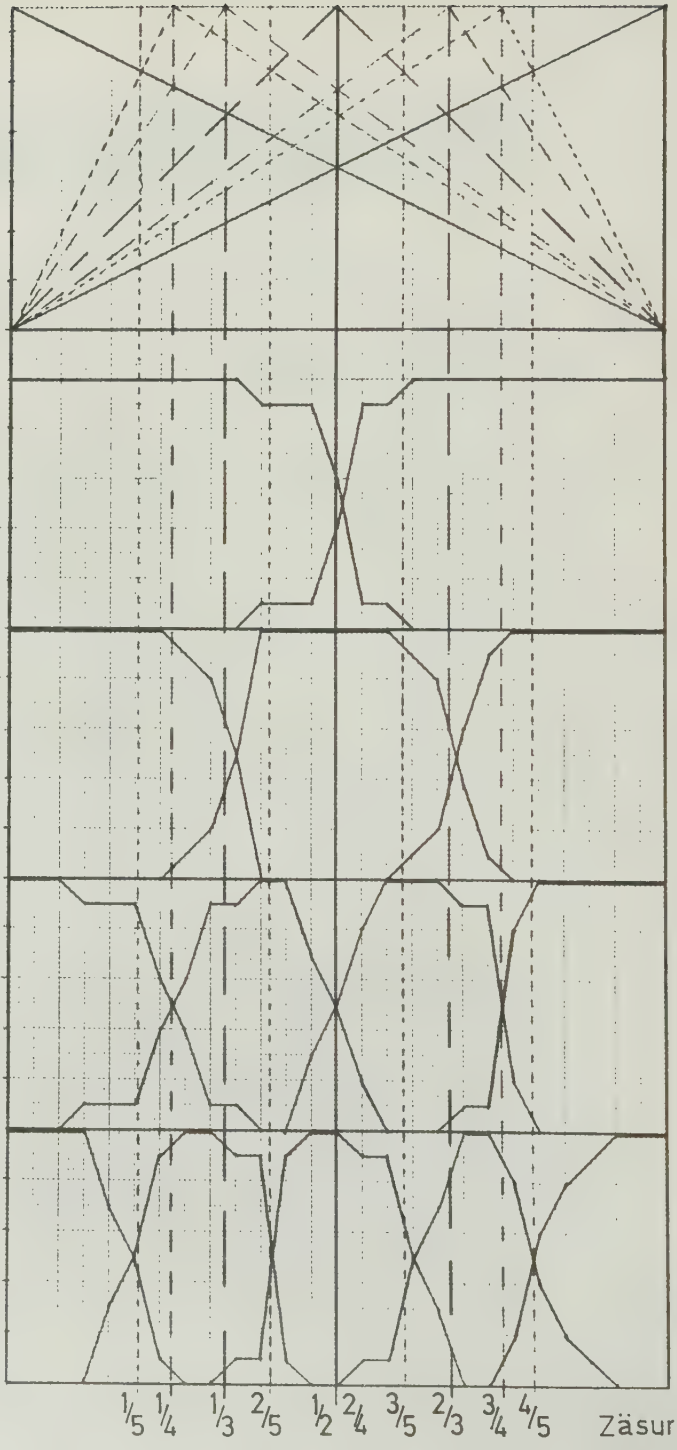


Abb. 2

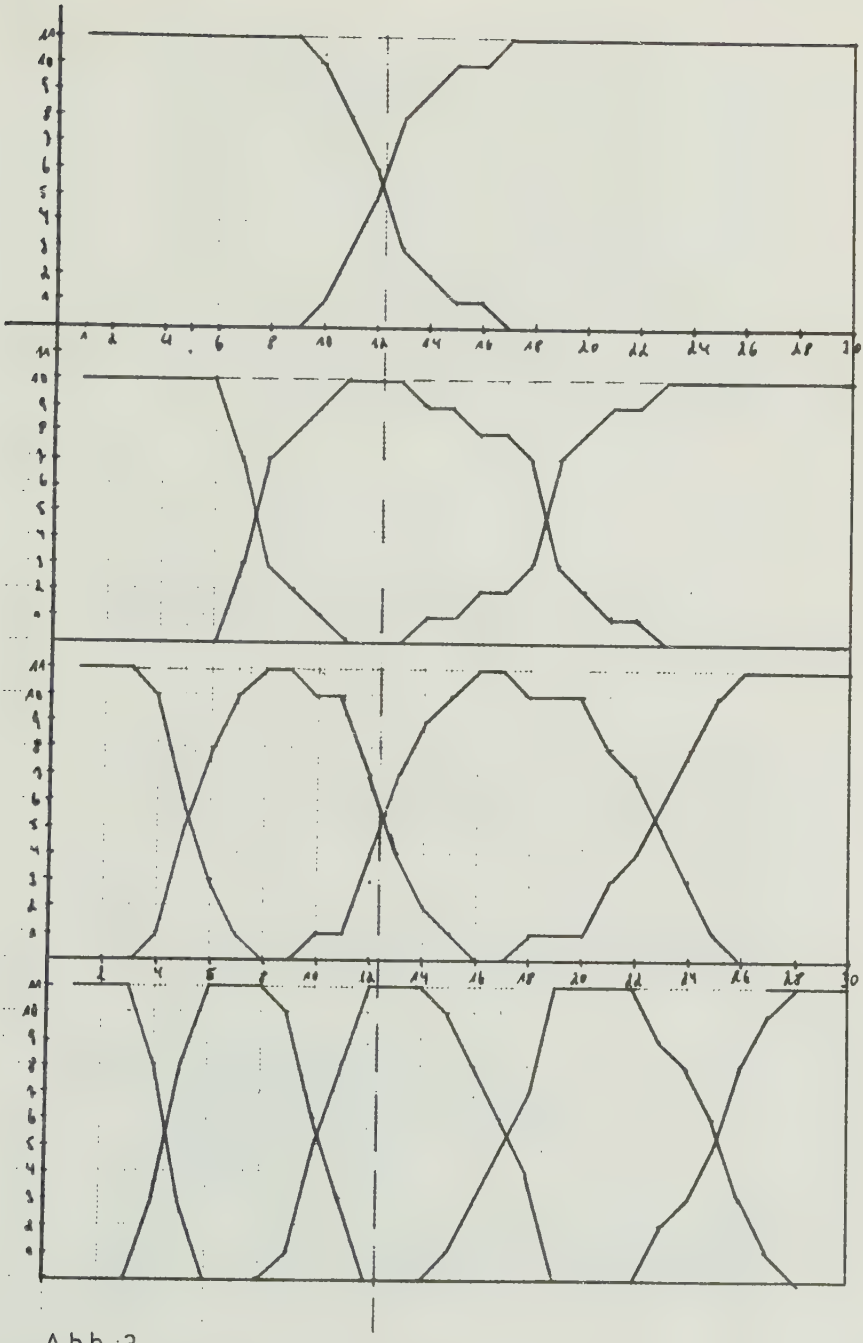


Abb. 3

Werden die Elemente der Trägermenge bezüglich des systemkonstituierenden Merkmals über eine physikalische Metrik gleichabständig angeordnet, dann kann die Bildmenge die Trägermenge in gleich große Teilmengen zerlegen, sie muß es aber nicht. Ein Beispiel hierfür zeigt Abb. 3 (über die Details des Experiments s. Anhang II). Das von WITTE angegebene Modell ist ein Ähnlichkeitsmodell und kein psycho-physisches. Die Linearitätsannahmen gelten für Ähnlichkeiten und nicht für psycho-physische Zuordnungen.

4. Einige Überlegungen zur Verteilung kategorialer Urteile.

Verteilen sich die individuellen Zäsurstellen zwischen je zwei Kategorien in irgendeiner Form (im allgemeinen annähernd normal, s. WITTE (33), ORLIK (37)) um eine Stelle im Kontinuum eines jeweiligen Merkmals, dann stellen die die zu einer gegebenen Zäsur benachbarten kategorialen Bereiche berandenden Kurvenzüge Verteilungsfunktionen dar

mit $F_1(x) = P(X \leqq x)$

und $F_2(x) = 1-F_1(x) = P(X \geqq x)$

wobei X gleich die betrachtete Zufallsveränderliche;
x gleich der jeweilige Wert, den X annimmt;
P gleich die Wahrscheinlichkeit für alle Werte $X \leqq x$ bzw. $X \geqq x$.

Der Sachverhalt der Verteilungsfunktion ist einleuchtend für die beiden Randbereiche; die beran-

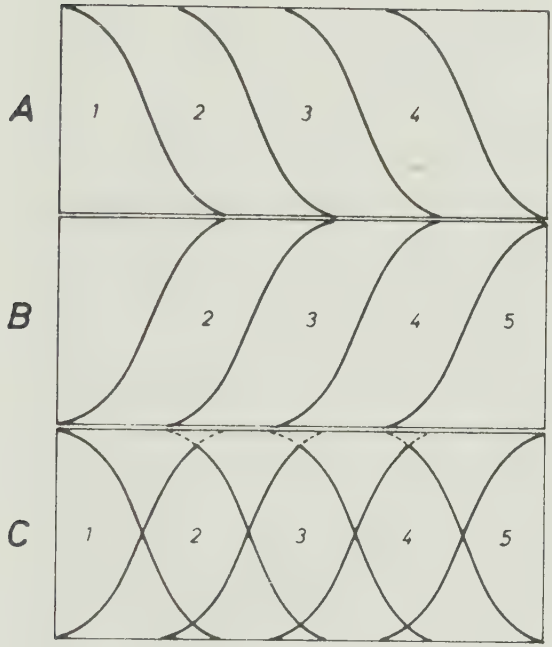


Abb. 4. Schema einer fünf-kategorialen Beurteilung.

A gleich links aufkumulierende Urteilsverteilungen.

B gleich rechts aufkumulierende Urteilsverteilungen.

C gleich Überlagerung von A und B.

Die Zahlen 1 bis 5 für irgendwelche geordnete Eigenschaften eines Merkmalskontinuums.

denden Kurven sind offensichtlich Summen-Kurven. (Die zu einer Summen-Kurve gehörige Summenfunktion bezeichnen wir auch - wie üblich - als eine Verteilungsfunktion.) Bei einem zwei-kategorialen System stehen die durch die beiden Kategorien gegebenen Verteilungen der kategorialen Urteile spiegelbildlich zueinander, also im obigen Verhältnis: $F_2(x) = 1-F_1(x)$. Dieser Sachverhalt ist schon aus dem üblichen Diagramm der Häufigkeitspolygone der kategorialen Urteile zu entnehmen. Für den Fall von drei- und mehr-kategorialen Systemen führt das Bild der üblichen Häufigkeitspolygone der kategorialen Urteile bezüglich der Frage, ob wahrscheinlichkeitsfunktionen oder Verteilungsfunktionen für die Verteilung der kategorialen Urteile anzunehmen sind (natürlich unter der Voraussetzung, daß die Menge der Urteile auf 1 normiert wird), zunächst in die Irre. Man kann aber eine Aufklärung erhalten, wenn man etwa von "links" nach "rechts" gehend den durch die den linken Randbereich berandende Kurve abgesteckten Sektor des Kontinuums, dann den rechts benachbarten entsprechenden Sektor usw. (also bei n Kategorien n-1 derartige Sektoren) betrachtet.

Die in den Sektoren zu findenden Paare bezüglich P komplementärer Verteilungsfunktionen entsprechen dann offensichtlich denen im Falle zweier Kategorien. Da jeweils zwei Funktionen zueinander komplementär sind, reicht zur Beschreibung der Verteilung der kategorialen Urteile bei n Kategorien eine Schar von n-1 Summenfunktionen aus. Die Schar kann dann entweder rechts oder links aufsummierend sein (s.Abb. 4A,B). Zeichnet man beide Scharen übereinander, ergibt

sich das übliche Bild der Häufigkeitsverteilungen kategorialer Urteile. In Abhängigkeit von der Streuung der Urteile überschneiden sich die Kurven mehr oder weniger stark (Abb. 4C), sodaß die Summenkurven nicht maximal ansteigen.

Die die mittleren Bereiche berandenden Kurven ergeben ein Bild, das so aussieht wie das einer Dichtefunktion (der Normalverteilung). Im allgemeinen wird es wohl auch gelingen, eine Dichtefunktion zu finden, die ausreichend gut an die empirische Urteilsverteilung angepaßt ist (wie dieses HRUSCHKA (14) versucht hat, übrigens nicht nur für die Verteilung von kategorialen Urteilen, sondern auch für die interindividuelle Verteilung der Lage von Systemen; sie spricht zwar nur von Normalverteilung, meint aber offensichtlich ihre Dichtefunktion). Die Tatsache einer ausreichenden Anpassung einer Dichtefunktion an die gegebenen Daten ist aber kein Beweis für die Frage nach Verteilungsfunktion oder Dichtefunktion. Das Problem muß theoretisch entschieden werden.

Andererseits lassen sich jedoch auch Argumente vorbringen, die für eine Dichtefunktion sprechen. Die Urteilsverteilungen der Randbereiche stellen dann "halbierte" Dichtefunktionen dar. Nimmt man aber an, daß es sich hier um Verteilungsfunktionen handelt, dann erscheint es notwendig, die Urteilsverteilungen als komponierte Verteilungsfunktionen wie oben beschrieben anzusehen.

5. Auswahl des Gegenstands-Systems und des experimentellen Vorgehens.

(a) Spezifität der Trägermenge.

Kehren wir zum Ausgangspunkt unserer Überlegungen im 1.Kapitel zurück. Wie wir dort gezeigt haben, ist die Frage nach Gütemaßstäben identisch mit sonstigen Fragen nach Bezugsystemen über sonstige Trägermengen systemkonstituierender Merkmale. Wenn es sich bei Untersuchungen von Bezugsystemen um die Erkundung der Gesetzmäßigkeiten der Zuordnung von Mannigfaltigkeiten hinsichtlich systemkonstituierender Merkmale zu System-Eigenschaften handelt, so bei unserer Fragestellung speziell um die Zuordnung von Leistungen zu (unscharf gesprochen) Leistungen charakterisierenden Eigenschaften.

Dieser Sachverhalt ist nun aber selbst nicht unproblematisch. Handelt es sich bei der Trägermenge um Striche, Gewichte, Quadrate, Kreise oder andere "abstrakte" Gegenstände, so unterscheiden sich die Elemente der jeweiligen Mengen im wesentlichen nur bezüglich der jeweils konstruierten Ausprägungsgrade des gewählten Merkmals. Also z.B. die Elemente einer Menge von Strichen nur hinsichtlich ihrer Länge, sodaß jeweils Gegenstand und Ausprägungsgrad gewissermaßen identisch sind. Verdeckt wird hierbei eine mögliche Unterscheidung, ob der Strich (also irgendein "Ding") bezüglich seiner Länge beurteilt wird oder eine Länge bezüglich ihrer "Langheit".

Solche Überlegungen sehen nun zunächst nach müßiger Sophisterei aus, führen aber ausgedacht zu einem Kernproblem aller Bezugsystemforschung, wichtiger und vorgängiger, wie uns scheint, als andere im Grunde erst nachzuschaltende Probleme etwa der Trennschärfe der kategorialen Bereiche (also des Grades der Differenziertheit eines Systems) und der meßtheotheoretischen Aspekte von Bezugsystemen.

Das Problem, das hier angesprochen werden soll, sei schlagwortartig mit dem Titel der "Spezifität der Gegenstandsmengen", über die sich Bezugsysteme ausbilden bzw. ausgebildet haben, bezeichnet. Die Frage hiernach muß jeweils gelöst werden, bevor sinnvollerweise über die "Struktur" des jeweiligen Systems gesprochen werden kann. Es dürfte allerdings nicht selten sein, daß das bezeichnete Problem in gegebenen Fällen methodisch vernünftig gar nicht angehbar ist, etwa dann, wenn der Beurteiler zwar die Trägermenge kennt, einzelne Elemente aus ihr aber nicht benennen kann, ohne daß eine Möglichkeit besteht, sie zu konstruieren (wie etwa bei Strichen).

Das Problem, das wir am Beispiel der Striche skizziert haben, stellt sich bei der hiesigen Fragestellung folgendermaßen dar: Was soll beurteilt werden? Die abstrakte Leistung, die unter Umständen in Metern, Kilogramm, Sekunden gemessen werden kann wie z.B. bei sportlichen Leistungen? Sind 56m weit oder nicht weit, gut oder schlecht? Die Frage ist in dieser Form offensichtlich absurd und kann nicht beantwortet werden. Beim Kugelstoßen sind 56m unerreichbar, beim Hammerwerfen eine

respektable, beim Speerwerfen eine schwache Leistung. Für den Fall des Kugelstoßens läßt sich das Urteil zwar ohne Einschränkung aufrechterhalten. In den anderen beiden Fällen erweist es sich aber, daß die Leistungen "gemessen" sind an den Weiltklasseleistungen. Aber auch nicht an Weltklasseleistungen schlechthin, sondern (wie man sich leicht anhand historischer Tabellen über Weltrekorde in den verschiedensten Sportarten leicht informieren kann) an denen vom Jahre 1968. Weiterhin muß möglicherweise spezifiziert werden daraufhin, ob es sich um die Leistung eines Professionellen oder eines Amateurs handelt; eines erwachsenen Hochleistungssportlers oder eines Nachwuchsmannes; eines Jugendlichen oder eines im, für die verschiedenen Sportarten verschiedenen, optimalen Leistungsalters Stehenden oder eines "alten Herrn".

Dieses Beispiel soll zeigen, daß es nicht möglich ist, nur "abstrakte" Leistungen beurteilen zu lassen, sondern daß spezifische Personengruppen (die Trägermengen) ausgewählt werden müssen, deren Mitglieder (die Elemente der Trägermenge) dann bezüglich ihrer Leistungen eingestuft werden. Die Gesamtheit der Mitglieder der durch einen Merkmalskomplex zusammengefaßten Gruppe wäre dann die Bezugsgruppe für über sie aufgespannte Systeme.

Der durch einen spezifischen Merkmalskomplex charakterisierte jeweilige systemkonstituierende Gegenstandsbereich ist es, den wir mit der Spezifität von Bezugsmengen gemeint haben. Es ist nicht auszuschließen, daß gegebenenfalls ein Merkmalskomplex ein System so spezifiziert, daß es seinen Geltungsbereich nur innerhalb eines "Umgebungsindividuums" besitzt (dieses also zum Komplex

gehört). Das hat zur Folge, daß das System umfunktioniert werden muß über dieselbe Gegenstandsmenge bei veränderten Umgebungsbedingungen.

Es ergibt sich damit die Notwendigkeit einer "Ökologie", besonders dann, wenn im Alltag stabilisierte Systeme unter Laboratoriumsbedingungen untersucht werden. Realiter dürfte die Forderung nach vorgängiger "Umgebungsforschung" zur Erkundung der spezifischen Einbettung von Bezugsystemen freilich meist nur approximativ möglich sein, in vielen Fällen nicht durchführbar mit entsprechenden Konsequenzen für nachgeschaltete strukturelle Überlegungen. Diese Umgebungsforschung muß allerdings eigenständig vorgenommen werden und kann selbst nicht wieder bloß gewissermaßen als Nebenprodukt des zur Erkundung der Struktur eines Systems aufgestellten Kategorisierungsexperimentes abgehandelt werden.

(b) Experimentelles Vorgehen.

Erkundung von Gütemaßstäben führt über Erkundung von Leistungsbeurteilungen. Leistungsbeurteilung ist Beurteilung von Mitgliedern einer Gruppe (der Bezugsgruppe) bezüglich ihrer Fähigkeit in einem gegebenen Leistungsbereich. (Der Begriff der Bezugsgruppe ist hier formal gemeint wie allgemein in der Bezugsystemforschung der einer Bezugsmenge, über die sich hinsichtlich eines Merkmals ein System aufspannt.) Bleiben wir mit unserer Darstellung im Rahmen der Leistungsmotivation, so ist die Bezugsgruppe "relevante" Bezugsgruppe, d.h. für die eigene Leistungseinstufung und eigene Leistungsanforderung relevante

Gruppe, d.h. die "leistungszugehörige" Gruppe (der Begriff wurde gewählt, um an dieser Stelle einer Diskussion über Mitgliedsgruppe-Bezugsgruppe und über mögliche Divergenzen beider aus dem Wege zu gehen).

Die Leistungsbeurteilungen sind also in diesem Zusammenhang "Binnen"-beurteilungen. Im Unterschied etwa zu Leistungsbeurteilungen von Lehrern, Meistern, Trainern über ihre Zöglinge, die für ihre eigenen Leistungsanforderungen im allgemeinen keine Mitglieder "relevanter Bezugsgruppen" darstellen.

Da es aus erhebungstechnischen Gründen bei Beurteilungen von Personen praktisch unvermeidlich ist, daß die Beurteiler die Beurteilten namentlich kennen, konnte die Datenerhebung nur über Mitgliedsgruppen(z.B. Schulklassen, Trainingsgruppen bei Sportvereinen) vorgenommen werden. Es schien aber nicht von vornherein ausgeschlossen, daß besonders bei Schulklassen durch das jahrelange Zusammensein der Mitglieder und durch die Tatsache, daß die Mitglieder in durch den Unterricht bedingten mannigfachen Konkurrenzen stehen, die Mitgliedsgruppe gleichzeitig auch relevante Bezugsgruppe ist. Es ist aber trotzdem daran festzuhalten, daß auch Schulklassen zunächst nur als Mitgliedsgruppen zu interpretieren sind.- Erheblicher nun als der erhebungstechnische Grund der Notwendigkeit der Benennung ist derjenige, daß die Mitglieder der Gruppe sich bezüglich der jeweiligen Leistung ausreichend kennen, sodaß eine Beurteilung überhaupt möglich wird.

Die gewählte Methode für unsere Datenerhebung besteht also darin, daß in geschlossenen Gruppen jeder jeden beurteilt (sich selbst ausgenommen) bezüglich eines Leistungsbereichs nach vorgegebenen Kategorien. Diese "Methode" ist nun nicht eine, die auch unter Umständen durch eine andere ersetzt werden könnte, sondern eine sachimmanente Notwendigkeit, wenn man die gefundenen Strukturen als Gütemaßstäbe im Rahmen der Leistungsmotivation interpretieren will. Mit der gewählten Methode ist also eine inhaltliche Entscheidung getroffen.

> Freilich: Verdacht auf Eingleisigkeit der gegebenen Interpretation von Gütemaßstäben zur Herstellung einer Notwendigkeit einer ausgewählten Art der Datenerhebung aus dem Begründungskontext Gütemaßstab: Bezugsgruppe: Gruppenmaßstab: Binnenbeurteilung ist nicht indiskutabel. Denn wenn auch Leistungsbeurteilungen nur innerhalb eines Bezugrahmens möglich sind, muß der Bezugrahmen Bezugsgruppe sein? Sind nicht "einsame" Leistungsanforderungen möglich, das "Bestmögliche" zu erreichen, wenn das Beste nicht nur das ist, was durch eine gerade gegebene Leistungsrangverteilung geliefert wird, sondern was von der Sachmöglichkeit her unüberbietbar ist, sodaß die Distanz zu diesem Besten über den Grad des Leistungsniveaus entscheidet. Ist dann eine Bezugsgruppe nötig?

(c) Auswahl der Merkmalskontinua. Die Untersuchungen von KELLNER.

Welche Leistungsbereiche sollen nun gewählt werden? Es schien sinnvoll, die Wahl zunächst nach folgenden Kriterien zu treffen:

1. Die Leistung soll "objektivierbar" sein, d.h. ganz naiv, daß sie direkt wahrnehmungsmäßig beobachtbar ist.

2. Sie soll allerdings ausreichend komplex sein, sodaß sie nicht entlang eines physikalischen Kontinuums abstufbar ist (wie etwa beim Kugelstoßen), und das Unternehmen hier wieder nur eine Variante der bisherigen psycho-physischen Untersuchungen über Bezugsysteme darstellen würde.
3. Die Leistungen müssen für die Untersuchten Gruppen einen ausgeprägten Leistungscharakter haben, wenn, nach HECKHAUSEN, in den Leistungsbereichen Leistungsmotivation sich ausgebildet hat, in denen ein Gütemaßstab besteht.

Alle drei Bedingungen scheinen für Gruppen von Jugendlichen für sportliche Leistungen von Art des Handballs oder Fußballs erfüllt zu sein, also Leistungen mit ausgesprochenem Kampfcharakter.

KELLNER (17) hat schon in einer Untersuchung auf Anregung von WITTE hin dieses Verfahren der gegenseitigen Beurteilung von Mitgliedern einer Gruppe bezüglich sportlicher Leistungen verwendet.(Dieses Verfahren hat offensichtlich eine gewisse Ähnlichkeit mit dem des Soziogramms.) Er erhielt nur bei einer Gruppe ein klares, für ihn interpretierbares Ergebnis, in der die Gruppenmitglieder sich nach 2,3,4,5 Kategorien hinsichtlich ihrer Fähigkeit im Fußballspielen beurteilten. Die Kategorien waren bei:

2 Kategorien: gut, schlecht;
3 " : gut, mittel, schlecht;
4 " : gut, überdurchschnittlich, unterdurchschnittlich, schlecht;
5 " : sehr gut, gut, mittel, schlecht, sehr schlecht.

Alle anderen vorgegebenen Eigenschaften versagten, klare Ergebnisse zu liefern. Allerdings versagte auch in diesem Falle die Kategorisierung nach

vier Kategorien: offensichtlich an der Unangemessenheit der beiden mittleren Eigenschaften.- Innerhalb der KELLNERschen Arbeit muß dieses singuläre positive Ergebnis (nach der Menge der nicht brauchbaren Ergebnisse) mehr oder weniger als Glückstreffer angesehen werden.

Wir wiederholten zunächst serienweise die KELLNERschen Versuche, um zu sehen, ob sich sein Ergebnis reproduzieren läßt. Dabei versuchten wir geignetere Eigenschaften-Systeme zu finden speziell für die Kategorisierung nach vier Kategorien, allerdings unter Beibehaltung der Eigenschaften-Systeme, die sich bei KELLNER als günstig herauskristallisiert hatten. Weiterhin wurde eine Erweiterung der Eigenschaften-Systeme auf sechs und sieben Eigenschaften hin versucht.

> In anderen, älteren, Arbeiten über Bezugsysteme, z.B. HRUSCHKA (14) war es bei psycho-physischen Systemen gelungen, bis zu sieben Kategorien hin trennscharfe Gliederungen zu erhalten. Sieben Kategorien scheinen allerdings das Maximum zu sein, bei dem noch Trennschärfe zu erreichen ist. S. hierzu auch G.A.MILLER (18).

Außerdem werden wir versuchen, eine breitere Basis zur Interpretation der Ergebnisse zu gewinnen als KELLNER sie entwickelt hatte.

Die oben (S. 35 f.) angegebenen Bedingungen ließen es ratsam erscheinen, Trainingsgruppen in Sportvereinen als Versuchsgruppen auszuwählen. Ein renommierter Münsteraner Sport-Club erklärte sich bereit, in die Namenslisten der Fußball-Jugendabteilung Einsicht nehmen zu lassen. Die drei zur Verfügung stehenden Trainingsgruppen hatten Stärken von über 30 bzw. 40 Mitgliedern, also für die Untersuchungszwecke geeignete Größen. Leider zeigte es sich aber, daß die Daten nicht auswertbar waren, da, trotz

gegenteiliger vorheriger Versicherung der Trainer, die Jugendlichen nur einen je und je verschiedenen Teil ihrer Kollegen namentlich kannten.

Wir entschlossen uns daraufhin weitere Untersuchungen in Schulklassen durchzuführen, da hier gewährleistet war, daß die Mitglieder der Gruppen sich gegenseitig namentlich kannten. Auch hier wurde so verfahren, daß Namenslisten der Gruppenmitglieder ausgeteilt wurden, auf denen die Beurteilungen vorzunehmen waren. Es trat bei den Untersuchungen in den Schulen nun allerdings eine empfindliche Schwierigkeit dadurch auf, daß Rektoren und Lehrer, bei allem -mehr oder weniger zögerndem-Entgegenkommen, nicht bereit waren, mehr als eine Unterrichtsstunde zur Verfügung zu stellen, sodaß die Schüler sich sukzessive nach den verschiedenen Eigenschaften-Systemen beurteilen mußten (sc. immer über denselben Leistungsbereich und jedes System für sich, auf separaten Blättern). Dieser Sachverhalt ist bei der Interpretation der Daten zu berücksichtigen. Ob entscheidende wechselseitige Einflüsse bestehen, kann aber nicht von vornherein behauptet werden.

Die verwendeten Eigenschaften-Systeme sind folgende:

2	Kategorien	: gut, schlecht;
3	"	: gut, mittel, schlecht;
4	"	: gut, eher gut, eher schlecht, schlecht;
5	"	: sehr gut, gut, mittel, schlecht, sehr schlecht;
6	"	: sehr gut, gut, eher gut, eher schlecht, schlecht, sehr schlecht;
7	"	: sehr gut, gut, eher gut, mittel, eher schlecht, schlecht, sehr schlecht.

Die Eigenschaften-Systeme mit 2,3,5 Kategorien wie bei KELLNER, die mit 4,6,7 Kategorien sind neu gebildet.

Im gegebenen Argumentationskontext, vornehmlich wegen der Notwendigkeit einer "Binnenbeurteilung", war es erforderlich in Hinsicht auf Fragestellungen nach maximaler Differenzierungsfähigkeit und nach Metrik -auf welche wir später eingehen- , daß ein und dieselbe Gruppe von "Versuchspersonen", also die Mitglieder der jeweils befragten Gruppen (Klassen), alle Beurteilungen nach allen Eigenschaften-Systemen vornahmen (jeweils aufsteigend erst nach 2, dann nach 3, ..., dann nach 7 Kategorien). KELLNER versuchte dieser Schwierigkeit der Nichtunabhängigkeit der Messungen durch "Fraktionierungsversuche" (in seinem Sprachgebrauch) aus dem Wege zu gehen: Die betrachtete Gruppe wurde in zwei gleich umfangreiche Untergruppen aufgeteilt; jede Untergruppe beurteilte nach je zwei Eigenschaften-Systemen an zwei verschiedenen Tagen. Dieses Verfahren vermeidet die o.a. Schwierigkeit nun allerdings nur am ersten Tag bei den dann vorgegebenen Eigenschaften-Systemen. Die Messung wird aber kaum unabhängiger, wenn sie nicht am selben Tag, sondern einige Tage später vorgenommen wird durch dieselben Beurteiler. Im Gegenteil besteht jetzt darüber hinaus, dadurch daß man dieselben Beurteiler hat, die Möglichkeit, daß auch noch die intersubjektive Unabhängigkeit der Beurteiler verloren geht - einfach dadurch, daß die Beurteiler sich in der Zwischenzeit darüber unterhalten, wie sie sich eingestuft haben. Weiterhin wird nicht selten das Problem auftauchen, daß durch die Fraktionierung die Beurteilergruppen zu klein werden.- Neben diese beiden formalen Argumente gegen die Behauptung der Verbesserung der Messungen durch Fraktionierung der untersuchten Gruppen tritt nun für die hiesige Untersuchung ein mehr inhaltliches: KELLNER interessierte sich nur für die Frage, ob bei Beurteilungen hinsichtlich komplexer Merkmale Zerlegungen von zu Bezugsystemen gehörigen Trägermengen sich nachweisen lassen, die denen entsprechen, die bisher bei einfachen Eigenschaften nachweisbar waren. Bei unserem Interpretationskontext zur Leistungsmotivation hin scheint es notwendig zu sein, wenn die gefundenen Strukturen als Gütemaßstäbe interpretiert werden sollen und diese als Gruppenmaßstäbe, die Beurteilungen aller Mitglieder einer jeweiligen Gruppe zur Messung auf einer gegebenen Systemhöhe (welche gegeben ist durch die Mächtigkeit einer zum Eigenschaften-System eines Bezugsystems gehörenden Menge der Eigenschaften) heranzuziehen.

Wir ließen in 8 verschiedenen Klassen Kategorisierungen (und im übrigen auch Rangreihen, s. Kapitel 9) durchführen. Es handelt sich in allen Fällen um Realschulklassen von ausschließlich Jungen. Die folgende Tabelle gibt eine Übersicht über Mitgliedsstärke der Klassen (n), durchschnittliches Alter ($\bar{x}$), Streuung der Altersverteilung (s), Anzahl der Eigenschaften des jeweiligen Systems mit der größten Systemhöhe (nKat), Nr. der zugehörigen Abbildung und Tabelle.

Klasse	n	$\bar{x}$	s	nKat	Abb.Nr.	Tab.Nr.
I	38	11.5	0.86	7	5	1
II	37	12.7	0.92	7	6	2
III	42	13.4	1.02	7	7	3
IV	39	14.6	0.98	7	8	4
V	33	15.8	0.95	7	9	5
VI	33	15.6	0.93	5	11	9
VII	37	14.7	0.98	5	14	12
VIII	39	14.4	1.01	5	15	/

Die Ergebnisse der Klassen I - V haben wir zusammengefaßt betrachtet; die Klassen VI,VII,VIII stellen Spezialfälle dar (Einzelheiten später hierzu im Text). Die Schüler waren gefragt worden, welches Ballspiel sie am liebsten spielten (es war durchweg Fußball). Es wurde sichergestellt, daß dieser Sport auch im Unterricht betrieben wurde, so daß die Schüler sich gegenseitig in ihren Leistungen bekannt waren.

Bemerkungen zu den folgenden Abbildungen:
Es sind zu sehen die Verteilungen der Urteilshäufigkeiten à Kategorie. Links befindet sich der "gut"-Pol, rechts der "schlecht"-Pol. Die Abfolge der Kategorien wie oben (S.38) angegeben. Der Ort gleichhäufiger Beurteilung durch benachbarte Kategorien stellt die Zäsurstelle benachbarter kategorialer Bereiche dar. Auf der y-Achse ist die Anzahl der Beurteiler abgetragen, auf der x-Achse die beurteilten Mitglieder der jeweiligen Gruppe (Klasse). Die hier gleichabständig gewählte Anordnung der Urteilsgegenstände stellt eine Rangreihe dar. Die Ermittlung einer derartigen Rangreihe ist freilich nicht unproblematisch. Aus Gründen der Vergleichbarkeit der Lage der Zäsuren war es erforderlich, daß pro Klasse über die verschiedenen Systemhöhen hinweg jede beurteilte Person denselben Rangplatz erhielt. Die einfachste Vorgehensweise wäre es gewesen, aus einem höherstufigen System (etwa dem der 5-Kategorisierung) die Rangreihe

zu ermitteln, da hier, im Unterschied zur 2- und 3-Kategorisierung, gleiche Rangplatzbesetzung im allgemeinen vermieden werden kann. Man müßte jedoch hier eine hohe Urteilskonsistenz von Stufe zu Stufe verlangen. Es war deshalb vorteilhafter, einen mittleren Rangplatz zu berechnen aus den Systemen, die in den einzelnen Diagrammen jeweils aufgezeichnet wurden, allerdings unter Fortlassung der 2-Kategorisierung, da diese wegen der häufigen identischen Beurteilungen der Personen zur Gewinnung von diskreten Rangplätzen wenig beiträgt. Ein rechnerisch einfacher Weg wäre es gewesen, eine Summe aller zu einer Person abgegebenen Urteile zu bilden, und über diese Summen eine Rangabfolge zu erstellen. Die Summenbildung kann so geschehen, daß der pro Systemhöhe "besten" Kategorie die Zahl 1, der "schlechtesten" eine Zahl m zugeordnet wird. Also bei 2 Kategorien "gut" gleich 1, "schlecht" gleich 2; bei 3 Kategorien "gut" gleich1, "mittel" gleich 2, "schlecht" gleich 3 usw. Dieser Weg ist jedoch nicht gangbar, da identische Zahlen in Systemen von verschiedener Höhe verschiedene metrische (topologische) Bedeutung besitzen. Ein optimaler Weg schien es zu sein, pro Systemhöhe über die pro Person abgegebenen Urteile eine Rangabfolge zu bilden und schließlich über die Mittelung der verschiedenen Rangabfolgen eine endgültige Rangabfolge zu erstellen. Für einige Klassen hatten versuchsweise beide der (zuletzt) beschriebenen Rechnungsmodi durchgeführt. Es zeigte sich, daß sich mit beiden eine weitgehend ähnliche Rangabfolge erzielen ließ mit wenigen Ausnahmen, bei denen lediglich eine Vertauschung benachbarter Rangplätze erfolgte, welche Vertauschung allerdings in keinem Fall einen Einfluß auf die Berechnung der Zäsurwerte gehabt hätte.
In den Abb.n wurden die Häufigkeitsverteilungen der Urteile nur bis zu der Systemhöhe wiedergegeben, bei der noch eben eine ausreichende Bereichsdifferenzierung erkennbar war, d.h. genauer, wenn Zäsuren noch ablesbar waren. Die durchgezogene senkrechte Linie im Mittelbereich der Diagramme gibt den Ort der (vermutlichen) Systemmitte wieder, also den der Zäsur 1/2, 2/4 und gegebenenfalls 3/6.
Die Zeichen bedeuten: ++ sehr gut, + gut. += eher gut, = mittel, -= eher schlecht, - schlecht, -- sehr schlecht und kennzeichnen die entsprechenden Bereiche bzw. Kurvenverläufe der Häufigkeitsverteilungen der Urteile.

Auf die Berechnung der Skalenwerte für die Zäsuren und ihre Systemwerte in den nachfolgenden Tabellen sowie auf die einzelnen Bezeichnungen wird im nächsten Abschnitt eingegangen.

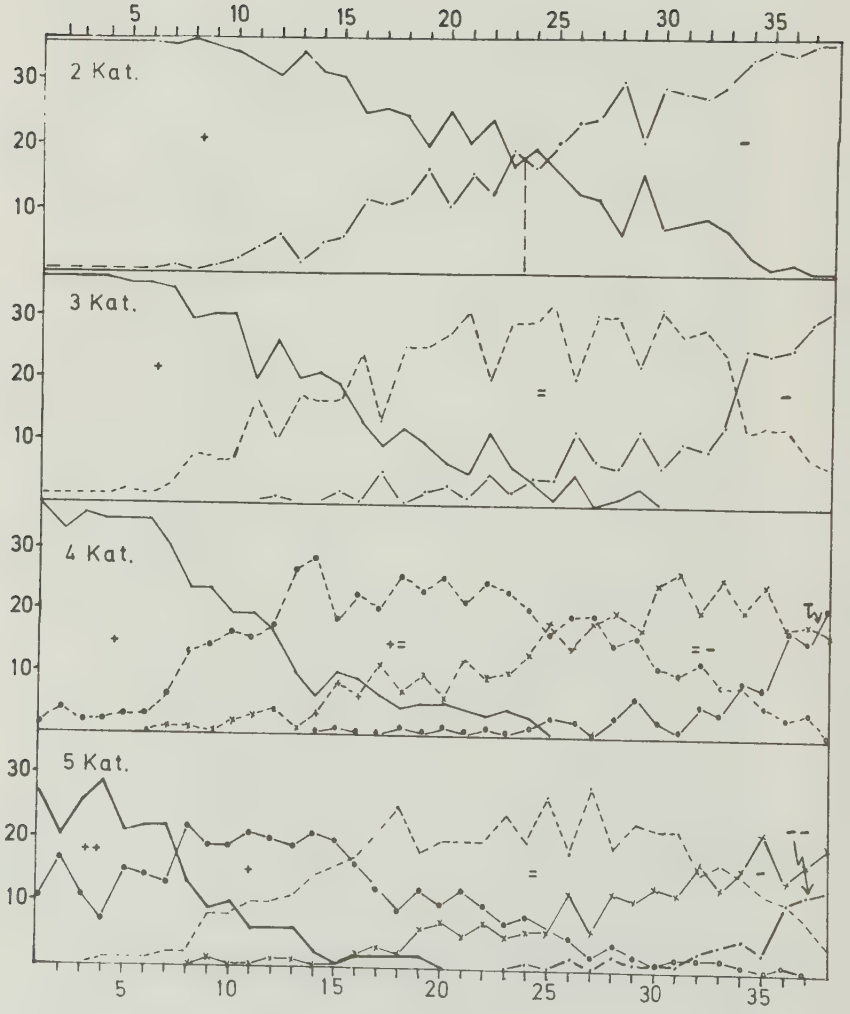


Abb. 5

Klasse I		Z-Werte				log-Werte			
		MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Lage d. Zäsuren nach d.empir.Skalenwerten									
Zäsur:									
1/2		1.269	0.257	0.257	0.257	2.274	0.445	0.445	0.445
1/3		0.756	-0.271	-0.271	-0.271	1.390	-0.463	-0.463	-0.463
2/3		1.919	0.877	0.925	0.936	3.446	1.563	1.664	1.692
1/4		0.435	-0.600	-0.600	-0.600	0.833	-1.035	-1.035	-1.035
2/4		1.479	0.472	0.472	0.472	2.664	0.846	0.846	0.846
3/4		2.121	0.935	1.140	1.183	3.760	1.621	2.049	2.164
1/5		0.175	-0.852	-0.872	-0.876	0.347	-1.507	-1.548	-1.557
2/5		0.928	-0.094	-0.094	-0.094	1.682	-0.163	-0.163	-0.163
3/5		1.749	0.749	0.749	0.749	3.150	1.345	1.345	1.345
Systemwerte:									
Zäsur:	erwartet:								
1/4	-25.0	-24.3	-24.0	-24.0	-24.0	-23.9	-23.8	-23.8	-23.8
3/4	25.0	25.5	19.0	24.5	25.5	23.4	18.0	24.5	25.9
1/5	-30.0	-30.4	-29.6	-30.1	-30.2	-30.3	-30.0	-31.3	-31.5
2/5	-10.0	-12.8	-13.8	-13.8	-13.8	-12.8	-12.8	-12.8	-12.8
3/5	10.0	10.7	11.4	10.2	10.0	10.3	10.6	10.1	9.8

Tab. 1. Zäsurwerte

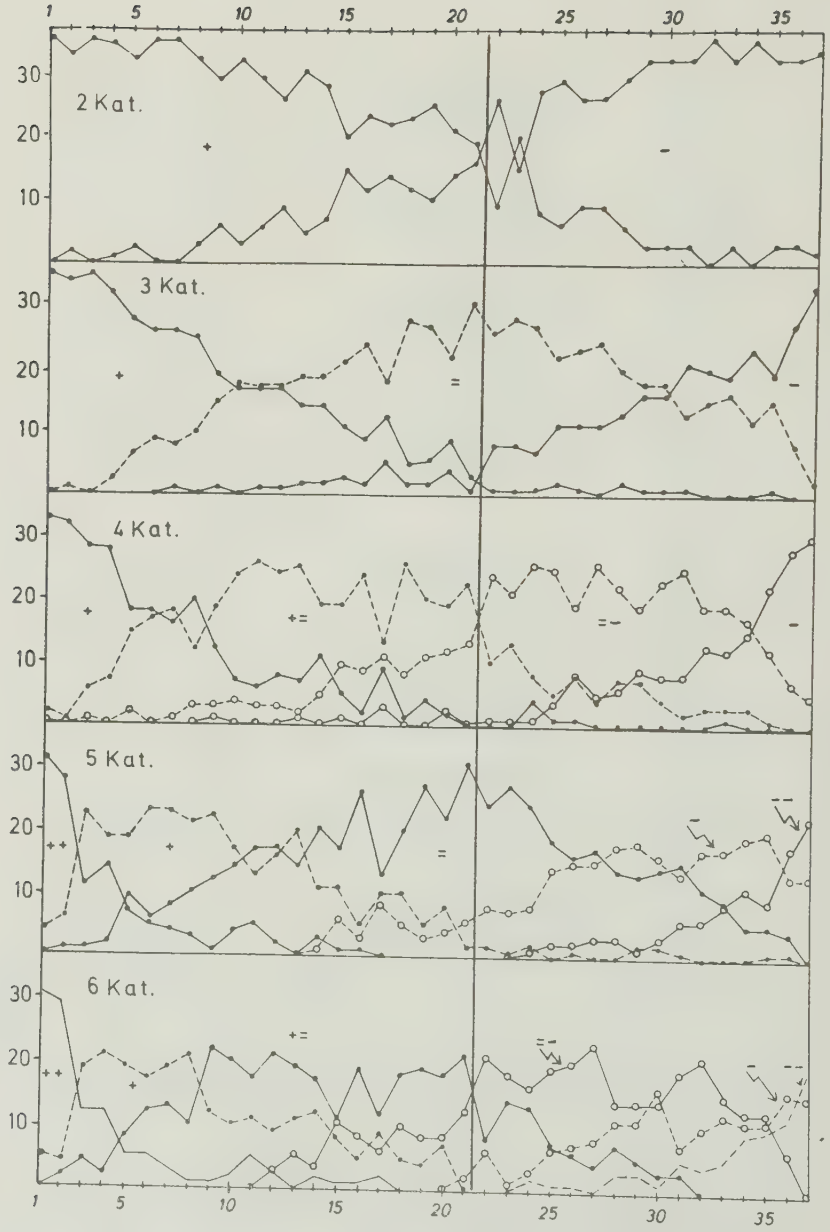


Abb. 6

Klasee II

Lage d. Zäsuren nach d.empir.Skalenwerten	Z-Werte				log-Werte			
	MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Zäsur								
1/2	1.462	0.256	0.256	0.256	2.861	0.459	0.459	0.459
1/3	0.724	-0.503	-0.503	-0.503	1.568	-0.870	-0.870	-0.870
2/3	1.948	0.755	0.755	0.755	3.697	1.319	1.319	1.319
1/4	0.403	-0.782	-0.782	-0.782	1.097	-1.354	-1.354	-1.354
2/4	1.462	0.256	0.256	0.256	2.861	0.459	0.459	0.459
3/4	2.022	0.773	0.830	0.840	3.817	1.337	1.454	1.493
1/5	-0.059	-1.194	-1.303	-1.331	0.113	-2.165	-2.391	-2.470
2/5	0.738	-0.488	-0.488	-0.488	1.595	-0.843	-0.843	-0.843
3/5	1.873	0.678	0.678	0.678	3.572	1.190	1.190	1.190
4/5	2.120	0.894	0.925	0.932	4.004	1.564	1.629	1.649
1/6	-0.059	-1.194	-1.303	-1.331	0.113	-2.165	-2.391	-2.470
2/6	0.565	-0.667	-0.667	-0.667	1.298	-1.152	-1.152	-1.152
3/6	1.462	0.256	0.256	0.256	2.861	0.459	0.459	0.459
Systemwerte:								
Zäsur: erwartet:								
1/4 -25.0	-23.9	-22.9	-22.8	-22.8	-22.8	-22.7	-22.7	-22.7
3/4 25.0	19.2	19.1	19.4	19.5	19.1	19.4	19.6	20.0
1/5 -30.0	-34.3	-31.8	-34.3	-34.8	-35.4	-33.0	-35.8	-36.9
2/5 -10.0	-16.3	-16.3	-16.3	-16.3	-16.3	-16.3	-16.3	-16.3
3/5 10.0	14.1	14.5	14.5	14.5	14.2	14.2	14.2	14.2
4/5 30.0	22.5	21.8	22.4	22.6	22.8	21.4	22.7	23.0
1/6 -33.3	-34.3	-31.8	-34.3	-34.8	-35.4	-33.0	-35.8	-36.9
2/6 -16.6	-20.0	-20.8	-20.8	-20.8	-20.1	-20.3	-20.3	-20.3
3/6 0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Tab. 2. Zäsurwerte.

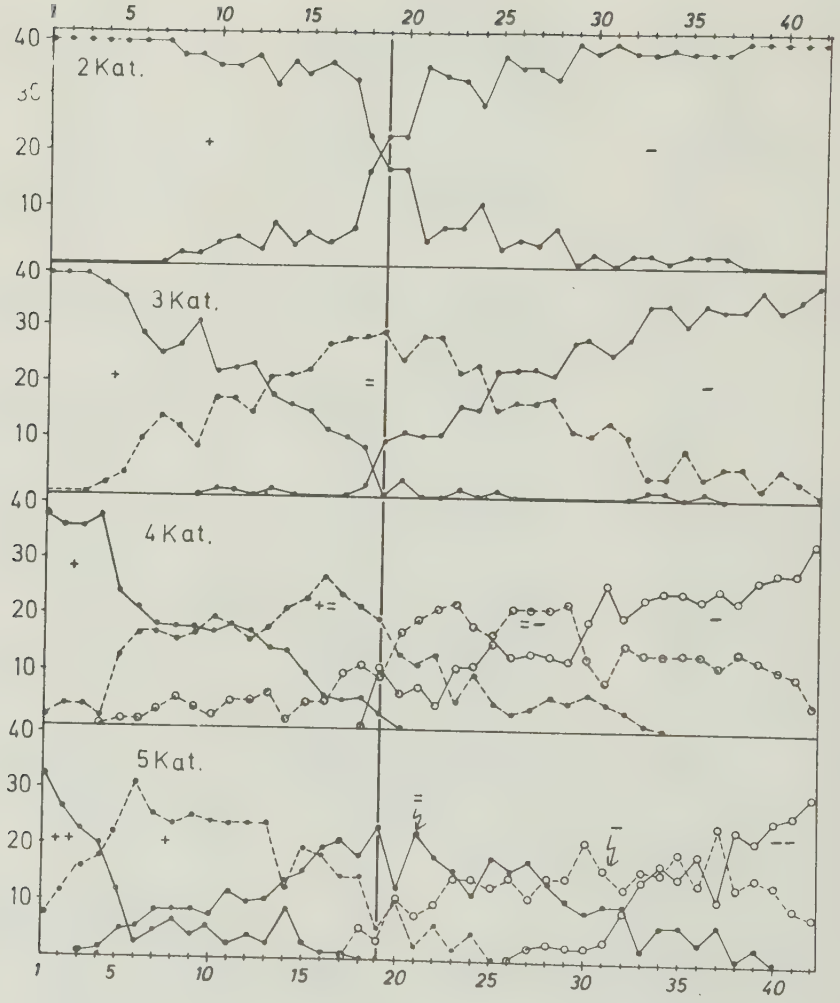


Abb. 7

Klasse III	Z-Werte				log-Werte			
	MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Lage d. Zäsuren nach d.empir.Skalenwerten								
1/2	1.027	-0.332	-0.332	-0.332	1.959	-0.574	-0.574	-0.574
1/3	0.592	-0.778	-0.778	-0.778	1.170	-1.383	-1.383	-1.383
2/3	1.845	0.450	0.521	0.532	3.424	0.825	0.964	0.989
1/4	0.264	-1.014	-1.126	-1.154	0.511	-1.877	-2.105	-2.173
2/4	1.541	0.139	0.216	0.227	2.878	0.271	0.421	0.447
3/4	2.136	0.656	0.810	0.838	3.924	1.167	1.476	1.545
1/5	-0.009	-1.239	-1.406	-1.448	-0.004	-2.308	-2.646	-2.749
2/5	0.717	-0.636	-0.654	-0.658	1.377	-1.144	-1.182	-1.191
Systemwerte:								
Zäsur: erwartet:								
1/4 -25.0	-22.7	-20.9	-22.5	-21.5	-23.0	-21.6	-23.3	-23.8
3/4 25.0	32.6	27.7	32.4	33.3	31.9	26.9	32.5	33.7
1/5 -30.0	-26.7	-25.0	-27.1	-27.1	-27.4	-25.9	-28.3	-28.1
2/5 -10.0	-14.4	-14.1	-13.6	-13.6	-14.6	-14.2	-14.8	-14.9

Tab. 3. Zäsurwerte.

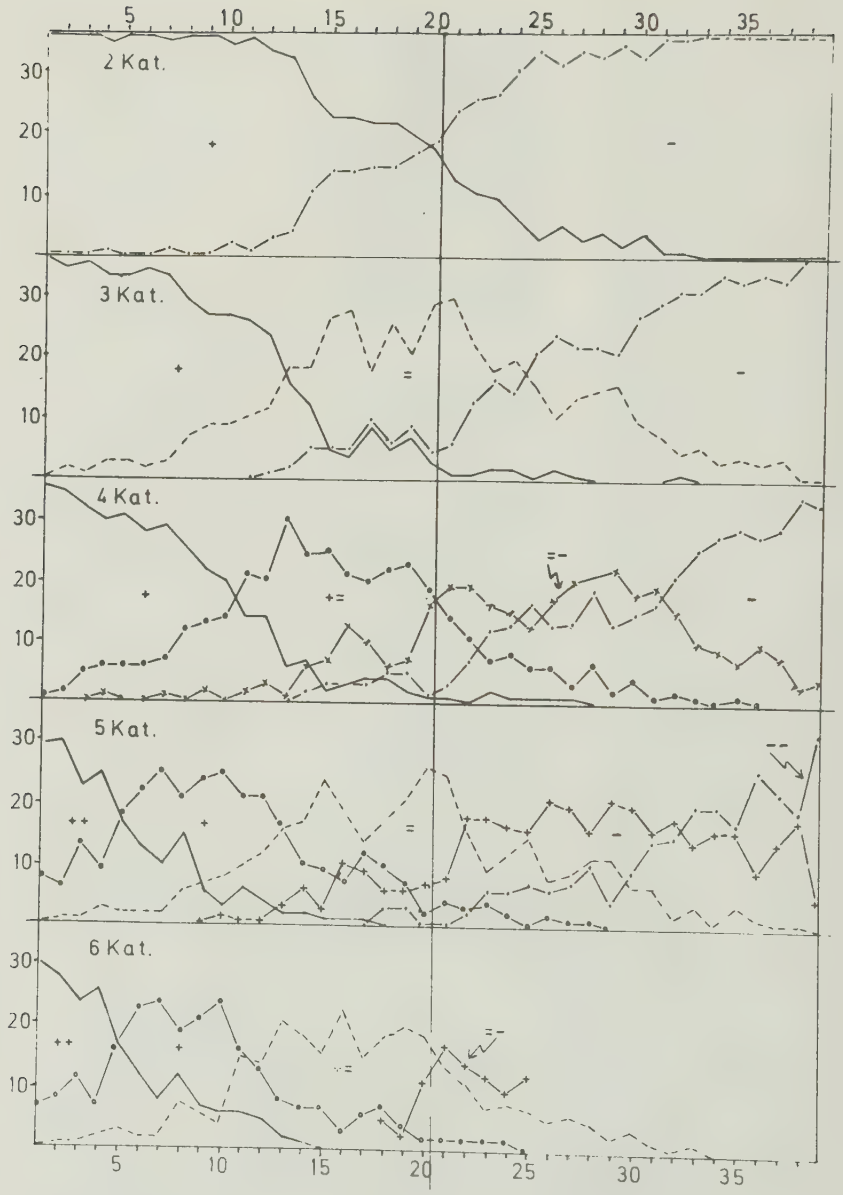


Abb. 8

Klasse IV	U	Z-Werte			log-Werte			
	MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Lage d. Zäsuren nach d.empir.Skalenwerten								
1Ø2	1.318	-0.028	-0.028	-0.028	2.432	-0.029	-0.029	-0.029
1/3	0.666	-0.689	-0.689	-0.689	1.234	-1.258	-1.258	-1.258
2/3	1.850	0.450	0.506	0.516	3.369	0.828	0.956	0.984
1/4	0.326	-0.930	-1.047	-1.072	0.616	-1.687	-1.923	-1.988
2/4	1.403	0.087	0.087	0.087	2.596	0.170	0.170	0.170
3/4	2.270	0.892	0.946	0.956	4.152	1.631	1.741	1.769
1/5	0.125	-1.122	-1.236	-1.263	0.251	-2.036	-2.271	-2.345
2/5	0.844	-0.515	-0.515	-0.515	1.569	-0.915	-0.915	-0.915
3/5	1.901	0.455	0.593	0.618	3.489	0.849	1.119	1.174
4/5	2.309	0.847	0.982	1.007	4.201	1.529	1.803	1.868
1/6	0.125	-1.122	-1.236	-1.263	0.251	-2.036	-2.271	-2.345
2/6	0.428	-0.892	-0.943	-0.954	0.797	-1.618	-1.721	-1.749
3/6	1.403	0.087	0.087	0.087	2.596	0.170	0.141	-0.141
Systemwerte: Zäsur: erwartet:								
1/4 -25.0	-27.5	-25.3	-28.2	-28.8	-27.5	-25.1	-28.3	-29.2
3/4 25.0	26.9	27.8	26.6	26.4	27.6	28.4	26.6	26.3
1/5 -30.0	-32.6	-30.0	-32.9	-29.9	-32.6	-30.0	-33.0	-34.1
2/5 -10.0	-14.3	-14.9	-14.9	-15.0	-14.2	-13.9	-14.7	▼14.7
3/5 10.0	15.6	12.8	15.8	16.1	15.7	13.2	16.0	16.5
4/5 30.0	28.4	23.4	27.9	28.2	28.5	26.4	27.6	27.9
1/6 -33.3	-32.6	-30.0	-32.9	-29.9	-32.6	-30.0	-33.0	-34.1
2/6 -16.6	-24.9	▼24.3	-25.6	-25.4	-25.0	▼20.0	-25.5	-26.0
3/6 0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Tab. 4. Zäsurwerte.

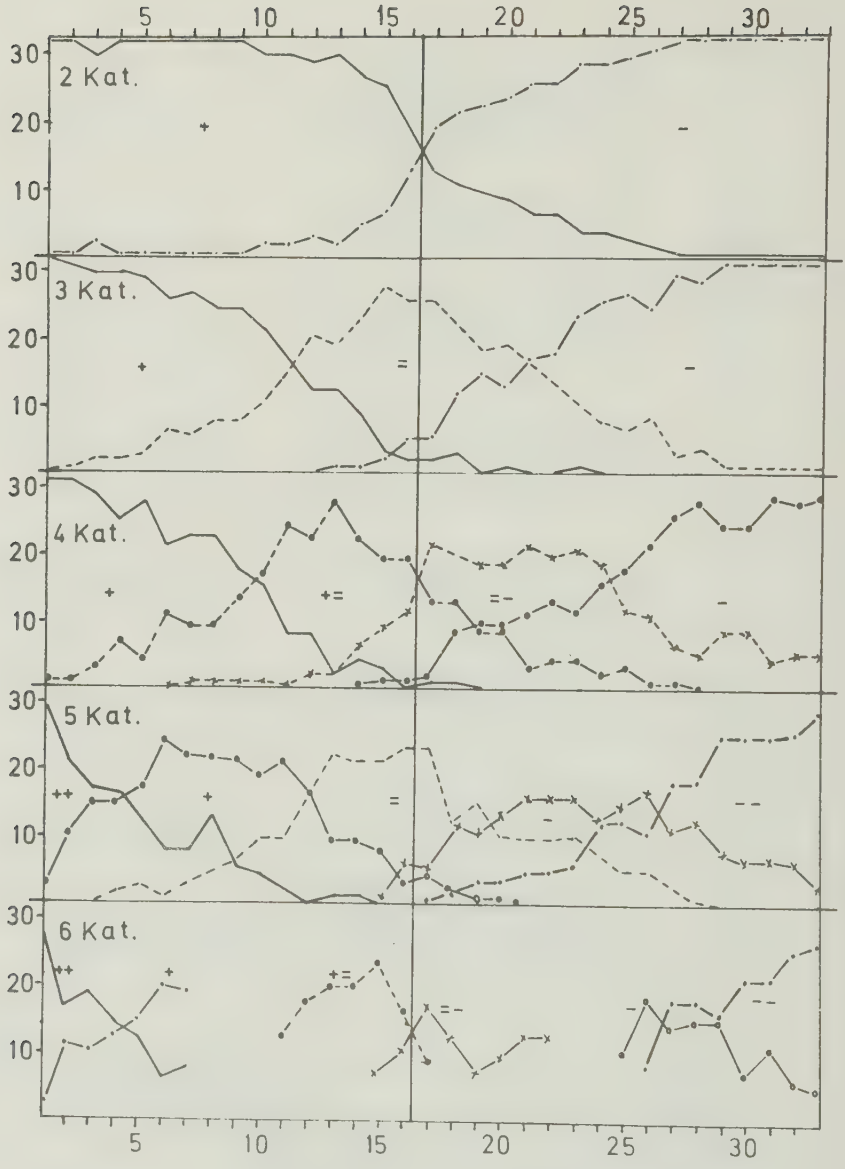


Abb. 9

Klasse V	Z-Werte				log-Werte			
	MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Lage d. Zäsuren nach d.empir.Skalenwerten								
1/2	1.357	0.052	0.052	0.052	2.471	0.109	0.109	0.109
1/3	0.592	-0.737	-0.737	-0.737	1.057	-1.351	-1.351	-1.351
2/3	1.852	0.515	0.565	0.572	3.346	0.925	1.023	1.046
1/4	0.335	-0.810	-0.995	-1.026	0.617	-1.459	-1.829	-1.928
2/4	1.357	0.052	0.052	0.052	2.471	0.109	0.109	0.109
3/4	2.233	0.725	0.944	0.977	4.036	1.297	1.731	1.842
1/5	0.002	-0.892	-1.281	-1.347	0.276	-1.596	-2.377	-2.598
2/5	0.732	-0.553	-0.602	-0.609	1.314	-1.014	-1.109	-1.132
3/5	1.731	0.392	0.444	0.451	3.137	0.712	0.814	1.030
4/5	2.344	0.723	1.039	1.073	4.226	1.294	1.896	2.051
1/6	0.002	-0.892	-1.281	-1.347	0.276	-1.596	-2.377	-2.598
2/6	0.489	-0.761	-0.845	-0.858	0.882	-1.383	-1.550	-1.594
3/6	1.357	0.052	0.052	0.052	2.471	0.109	0.109	0.109
5/6	2.344	0.723	1.039	1.073	4.226	1.294	1.896	2.051
Systemwerte:								
Zäsur: erwartet:								
1/4 -25.0	-22.3	-18.0	-22.1	-22.7	-22.0	-18.0	-22.1	-23.2
3/4 25.0	29.2	24.2	28.8	30.0	29.8	24.2	23.9	25.0
1/5 -30.0	-29.5	-20.0	-28.1	-27.5	-28.8	-20.0	-28.3	-30.8
2/5 -10.0	-13.6	-12.7	-14.8	-13.0	-13.5	-12.8	-13.8	-14.1
3/5 10.0	12.6	12.2	12.7	12.8	12.7	12.3	10.4	13.3
4/5 30.0	33.2	24.1	32.0	32.7	33.4	24.2	26.2	28.0
1/6 -33.3	-29.5	-20.0	-28.1	-27.5	-28.8	-20.0	-28.8	-30.8
2/6 -16.6	-18.7	-16.8	-18.8	-20.0	-18.7	-17.0	-18.9	-20.0
3/6 0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5/6 33.3	33.2	24.1	32.0	32.7	33.4	24.2	26.2	28.0

Tab.5. Zäsurenwerte.

6. Gewinnung einer Skala.

(a) Allgemeiner Begriff einer Skala.

Sollen die in den Abb.n 5 - 9 gefundenen Strukturen als Gütemaßstäbe interpretiert werden, dann, unter Beibehaltung der Interpretation, daß Gütemaßstäbe die Leistungen innerhalb einer Bezugsgruppe messen - durch welche Annahme wir den Zusammenhang von Gütemaßstäben und Bezugsystemen begründet hatten -, reicht die Bestimmung eines Bezugsystems als das oben angegebene geordnete Tripel allerdings nicht aus. Denn sprechen wir von Maßstäben, dann ist irgendeine Verwendung von Zahlen mitgemeint. Was ist nun Messen, bzw. welche Bedingungen muß es erfüllen? Nach PFANZAGL (22, S. 14) (das Zitat etwas länger, damit verständlich wird, wovon im "Grundsatz" die Rede ist):

> "Die folgenden Betrachtungen beziehen sich auf eine abstrakte Menge M. Wir werden später M interpretieren als die Menge aller möglichen Intensitäten des zu messenden Merkmals. Diese Intensitäten existieren natürlich nur an konkreten Objekten und alle Relationen, die zwischen diesen Intensitäten festgestellt werden, gehen letztlich auf Relationen zwischen konkreten Objekten zurück. (Man könnte sogar M selbst als die Menge aller Objekte interpretieren, die das betreffebde Merkmal in irgendeiner Intensität aufweisen.) Die allgemeinste Formulierung des von uns zu behandelnden Problems (sc. des Messens) ist durch folgenden Grundsatz gegeben:
> GRUNDSATZ: Die Menge M ist so auf die Menge der reellen Zahlen abzubilden, daß die Struktur der Bildmenge isomorph zur Struktur von M ist.
> Dadurch ist es möglich, aus der Struktur der Bildmenge Rückschlüsse auf die Struktur der abgebildeten Menge zu ziehen ..."

> Die hier von PFANZAGL vorgenommene Gleichsetzung der Menge der Intensitäten mit der Menge der Objekte ist allerdings unzulässig. Er würde hiermit voraussetzen, daß die Menge "aller möglichen" Intensitäten gleichmächtig ist der Menge der einschlägigen Objekte. Da der Fall nicht auszuschließen und im allgemeinen auch gegeben ist, daß die Menge der Objekte mächtiger ist als die der Intensitäten und, da es um die Messung der Intensitäten des betrachteten Merkmals geht, ist es nötig Äquivalenzklassen in der Menge der Objekte zu bilden bezüglich der Intensitäten, um Isomorphie der Abbildung zu erreichen.

Eine ähnliche, aber präzisere Definition einer Skala geben SUPPES & ZINNES (28, S. 11):

> "Let $\mathcal{A}$ be an empirical relational system, let $\mathcal{B}$ be a full numerical relational system, and let f be a function that maps $\mathcal{A}$ homomorphically onto a subsystem of $\mathcal{B}$ ... We say then that the ordered triple $\langle \mathcal{A}, \mathcal{B}, f \rangle$ is a <u>scale</u>."

(b) Abbildungsverhältnisse bei einer Kategorien-Skala.

Stellt ein Bezugsystem $\mathcal{S} = \langle \mathcal{T}, \mathcal{E}, \mathcal{Z} \rangle$ ein derartiges empirisches Relationen-System dar, dann wäre das geordnete Tripel $\langle \mathcal{S}, \mathcal{B}, f \rangle$ eine Skala. Man findet also bei Skalen über Bezugsysteme folgende Abbildungsverhältnisse vor:

f: Gegenstände $\longrightarrow$ Kategorien $\longrightarrow$ Zahlen .

Da das Eigenschaften-System $\mathcal{E}$ aus $\mathcal{S}$ geordnet ist, läßt sich den Kategorien unmittelbar eine beliebige Menge von Zahlen, die gleichmächtig ist der der Kategorien, zuordnen unter der Bedingung, daß die Ordnung der Zahlen die des Eigenschaften-Systems

abbildet und nichts als diese Ordnung abbildet.

Es seien:

(1) $\{a,b,c,\ldots,f,g,h,\ldots,k,l,m,\ldots,r,s,t,\ldots,x,y,z\}$ gleich Menge der Gegenstände, die Träger eines betrachteten Merkmals sind;

(2) $\{I,J,K,L,M\}$ gleich Menge der Kategorien, die der Bereich des Eigenschaften-Systems $\mathcal{E}$ ist;

(3) Z_1, Z_2, Z_3 gleich irgendwelche Mengen von Zahlen.

Dann wären derartige Skalen folgende:

Gegenstände	Abbildung	Kategorien	Abbildung	Zahlen Z_1	Z_2	Z_3
a b c . . f g h . . k l m . . r s t . . x y z	→	I	→	1	1	0.4
	→	J	→	2	4	1.7
	→	K	→	3	9	4.1
	→	L	→	4	16	4.8
	→	M	→	5	25	9.6

Bilden die Kategorien, in welcher Weise auch immer, die Gegenstände ab, so tun das auch alle drei Zahlensysteme bezüglich der Kategorien und damit auch

bezüglich der Gegenstände. Alle drei Zahlensysteme sind gleichberechtigt, da durch jedes von ihnen die Ordnung unter den Kategorien ablesbar ist. Es handelt sich hier also um Skalen vom Range einer Ordinal-Skala oder topologischen Skala (PFANZAGL), und jede Skala ist offensichtlich eindeutig nur bis auf monotone Transformation. $Z_1 \longrightarrow Z_2$ entspricht offensichtlich der Gleichung $y = x^2$; $Z_1 / Z_2 \longrightarrow Z_3$ offensichtlich monoton, Gleichung uns allerdings nicht bekannt.

Formal erfüllt zwar eine Ordinal-Skala die Bedingungen einer Skala, die Auskunft über die metrische Struktur der zugrundeliegenden Trägermenge ist aber gering. Im allgemeinen glauben wir auch, wenn vom Messen die Rede ist, eine Auskunft nicht nur über die Anordnung der Gegenstände erhalten zu dürfen, sondern auch über ihre Abstände. D.h. wir müssen eine zulässige Abbildung des Eigenschaften-Systems auf ein numerisches System finden, welche es erlaubt, nicht nur Anordnungen, sondern auch Distanzen abzulesen. Wie soll das aber erreicht werden, da die empirische Operation der Kategorisierung nur Aussagen über Anordnungen zuläßt.

Vor dieser Diskussion erscheint es freilich sinnvoll, zunächst zu prüfen, ob überhaupt die geforderten Ordnungsrelationen empirisch erfüllt sind. Tab. 6 zeigt, daß im allgemeinen, soweit die Zäsuren aus den Abb.n 5-9,11,14 überhaupt ablesbar waren, die Anordnung der Zäsuren den Erwartungen entspricht. Erwartungswidrig sind, von einer Ausnahme abgesehen, die Ordnungsrelationen zu den Zäsuren der 6-Kategorisierung, die freilich in keinem einzigen Fall für alle Kategorien trenn-

Zäsur	1/6	1/5	1/4	1/3	2/6	2/5	1/2	2/4	3/6	3/5	2/3	4/6	3/4	4/5	5/6
Klasse	(	(	(	=	(	(	=	=	(	(	=	(	(	(	
I	?	(	(	?	(	(	(!	?	(	(	?	(	?	?	
II	=!	(	(	)!	(	(	=	=	(	(	?	(	(	(	
III	?	(	(	?	(	(	=	?	?	(	?	(	?	?	
IV	=!	(	(	(!	(	(	=	=	(	(	?	(	(	?	
V	=!	(	(	=	(	(	=	=	(	(	?	(	(	=!	
VI	/	/	(	/	/	/	=	/	/	(	/	(	/	/	
VII	/	(	(	/	(	(	=	/	(	(	/	(	(	/	
VIII	/	?	(	/	=!	(	=	/	(	=!	/	(	?	/	

Tab. 6. Relationen der Zäsuren pro untersuchter Klasse.
1. Reihe: Namen der Zäsuren.
2. Reihe: Erwartete Relationen.
Die Zeichen bedeuten: " (" = kleiner als
") " = größer als
" = " = gleich groß wie
" ? " = nicht ablesbar
"R! " = erwartungswidrige Relation
" / " = entfällt (da das nächsthöhere System nicht abgefragt wurde)!

scharfe Bereiche lieferte. Die 7-Kategorisierung erbrachte keine ablesbaren Zäsuren und wurde deshalb auch schon in den entsprechenden Abbildungen nicht dargestellt.

Kehren wir nochmal zum Ausgangspunkt unserer Überlegungen zurück. Zur Frage der Gütemaßstäbe untersuchen wir Bezugsysteme. Hat sich ein Bezugsystem ausgebildet, dann ist die zugehörige Trägermenge abgeschlossen und hat bezüglich des betrachteten Merkmals ein maximales und ein minimales Element. Die Distanz zwischen diesen beiden Elementen läßt sich auf das geschlossene Intervall [0,1] normieren. Die Behauptung des WITTEschen Modells war dann, daß jedes polar geordnete, symmetrische Eigenschaften-System eine Zerlegung der Trägermenge derartig liefert, daß das Intervall [0,1] in gleich große Intervalle aufgeteilt wird. Nach WITTE (35) sind die metrisch relevanten Örter dieser Zerlegung die Grenzen der Teilintervalle (die Zäsuren). Das Modell macht dann Voraussagen für Klassen äquivalenter Zäsuren, die sich abbilden auf Klassen äquivalenter Paare ganzer Zahlen des Intervalls [0,1] der reellen Zahlen.

> Freilich würde für die Abbildung die Menge der rationalen Zahlen des Intervalls [0,1] ausreichen, da durch die Kategorisierung das Intervall in gleich große Teilintervalle zerlegt wird. Die gegebene Menge der rationalen Zahlen ist abgeschlossen gegenüber den einschlägigen Operationen, da man durch weitere Unterteilungen bzw. Zusammenlegungen immer wieder zu rationalen Zahlen des gegebenen Intervalls gelangt.

Diese Klassen äquivalenter Zäsuren lassen sich unmittelbar aus den Diagrammen der Verteilungen der Urteilshäufigkeiten ablesen (s. Abb.n 2,3, 5-9,10,11,14). Für die Äquivalenzklasse (1/2,2/4, 3/6) liegen die Zäsuren durchweg am selben Ort

(von unerheblichen Abweichungen abgesehen), ebenso für die Klasse (1/3,2/6) in Abb.9. Bei den anderen Zäsuren ist eine Bestätigung kaum zu erreichen, da diese Bestätigung der Zäsuren eines Systems von der Höhe n immer nur zu erreichen ist von Systemen, deren Höhe ein Mehrfaches von n betragen (also mindestens 2n). Im allgemeinen dürften aber die Zäsuren bei Systemen mit mehr als fünf Kategorien nicht mehr ablesbar sein, da die kategorialen Bereiche dann nicht mehr ausreichend trennscharf sind.

Wir haben der Kürze halber die Zäsuren mit den angegebenen rationalen Zahlen bezeichnet ohne hiermit etwas über den metrischen Ort ausgesagt haben zu wollen.

Im Falle der Äquivalenzklasse mit dem Repräsentanten 1/2 liegt die Vermutung allerdings nahe, daß es sich hier um die Mitte des Systems im metrischen Sinne handelt, d.h. daß die Distanzen vom Ort der Zäsur bis zu den beiden Endgliedern des Systems gleich groß sind. Aber selbst derartige gegenseitige Bestätigungen äquivalenter Zäsuren stellen noch keinen Beweis dar für die Metrizität der durch die Kategorisierung möglicherweise intendierten Distanzen. Hierfür ist ein von der Kategorisierung unabhängiges Meßverfahren erforderlich.

(c) Metrische Axiomen-Systeme.

i. PFANZAGLs Axiomen-System.

PFANZAGL liefert ein Axiomen-System, dessen Bisymmetrie-Axiom es erlaubt, durch ständige Mittelungen metrische Skalen zu erhalten.

> PFANZAGL faßt unter "metrische Skalen" alle Skalen, die es erlauben, über Distanzen Aussagen zu machen. Speziell diejenigen, die üblicherweise "Intervall"-Skalen genannt werden (also Skalen, die eindeutig sind bis auf lineare Transformation: $y = a + bx$), und diejenigen, die üblicherweise "Verhältnis"-Skalen genannt werden (also Skalen, die eindeutig sind bis auf Ähnlichkeits-Transformation bzw, bis auf "Streckungen": $y = ax$.)

Dieses Verfahren scheint geeignet zu sein speziell zur Bestätigung der durch die Kategorisierung vermuteten wichtigen System-Mitte.

> Darauf hat schon WITTE hingewiesen. In einigen älteren Arbeiten zur Theorie der Bezugsysteme - z.B. WEBER (32), AHLERT (1) - wurde das Verfahren nach PFANZAGL mit derselben Absicht wie hier nach Berichten mit Erfolg angewendet.

Wir versuchten diesen Weg zu gehen. Das Ergebnis war ein vollständiger Mißerfolg (s. Anhang III).

Wodurch ist dieser Mißerfolg bedingt? Lag es an den Beurteilern, die nicht imstande waren, über das betrachtete Merkmal Distanzen anzugeben? Es ist auch nicht ausgeschlossen, daß die Forderungen des PFANZAGLschen Axiomen-Systems inadäquat sind für die von uns herangezpgenen Gegenstandsmengen. Betrachten wir uns seine Axiome: PFANZAGL (22, S.20):

"Axiome der metrischen Verknüpfung:

1. E x i s t e n z - Axiom: Je zwei Elementen $a,b \in M$ ist eindeutig ein Element $a \circ b$ zugeordnet.
2. M o n o t o n i e - Axiom: Aus $b \; ((=)) \; b'$ folgt: $a \circ b \; ((=)) \; a \circ b'$ für alle a, aus $a \; ((=)) \; a'$ folgt: $a \circ b \; ((=)) \; a' \circ b$ für alle b.

3. S t e t i g k e i t s - Axiom: Die Mengen
$\{ x \mid M, x \circ b \gtrless a \}$ und
$\{ x \mid M, x \circ b \lessgtr a \}$ sind für alle a,b offen. Das Gleiche gilt für b o x.

4. B i s y m m e t r i e - Axiom:
$(a \circ b) \circ (c \circ d) \overline{=} (a \circ c) \circ (b \circ d)$.

.....

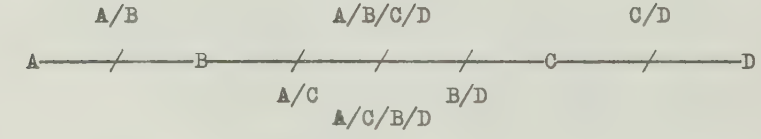
Wir bezeichnen eine metrische Verknüpfung eine Mitten-Bildung, wenn sie außer den Axiomen 1 - 4 noch die nachstehend angeführten Axiome erfüllt:

Axiome der Mitten-Bildung:

5. K o m m u t a t i v i t ä t s - Axiom:
$a \circ b \overline{=} b \circ a$.

6. R e f l e x i v i t ä t s - Axiom:
$a \circ a \overline{=} a$.

Nachstehende Abbildung veranschaulicht die Bedeutung der Bisymmetrie im Falle der Mitten-Bildung:



(Zitatende)

Die verwendeten Zeichen bedeuten:

$\lessgtr$	Elementbeziehung
$\gtrless$	kleiner,gleich,größer
o	allgemeines Verpnüpfungszeichen
$\{ \mid \}$	Klammer zur Bezeichnung von Mengen (an Stelle der üblichen geschweiften Klammern)
$\overline{=}$	Äquivalenzrelation

Das Bisymmetrie-Axiom, das die empirische Operation der Mitten-Bildung steuert, ist abhängig (nicht: ableitbar) vom Stetigkeits- wie auch vom Existenz-Axiom (natürlich auch vom Monotonie-Axiom, was uns aber jetzt nicht zu interessieren braucht). Das Existenz-Axiom selbst vom Stetigkeits-Axiom: denn ist die Menge nicht stetig, kann sinnvollerweise nicht die Eindeutigkeit einer Zuordnung eines dritten Elements zu zwei gegebenen Elementen einer Menge gefordert werden. Die Eindeutigkeit der Zuordnung wird aber für das Bisymmetrie-Axiom vorausgesetzt.

Für empirische Verhältnisse übersetzt fordern die Axiome 1 - 4 ein Kontinuum über irgendein Merkmal. Dieses Kontinuum ist in irgendeiner Hinsicht immer ansteigend. Zu je zwei beliebigen Schnittstellen auf diesem Kontinuum läßt sich eindeutig eine dritte zuordnen, die sich gegebenenfalls als Mitte zu jenen beiden auffassen läßt. Die empirisch gegebene Menge müßte zwar der Forderung des Stetigkeits-Axioms entsprechend nicht infinit sein. Sie könnte auch finit sein (jedenfalls in der Psychologie), sie müßte aber phänomenal dicht sein, d.h. phänomenal ohne Sprünge und Lücken. Das kann man im gegebenen Fall dadurch erreichen, daß in der Kette der Intensitäten (oder Ausprägungsgrade) des betrachteten Merkmals die Differenz benachbarter Elemente unterschwellig bleibt.(Daß man dennoch eine Kette erhält, liegt daran, daß die Indifferenz-Relation nicht transitiv ist - im Unterschied zur Äquivalenz-Relation.) Das verlangt aber recht umfangreiche Reizserien. Die Herstellung derartiger Serien ist aber nur möglich in psycho-physischen Untersuchungen.

Von diesen Überlegungen her ist es nun möglich, einige Interpretationsansätze dafür zu gewinnen, daß das Mittelungsexperiment bei uns mißlungen war. Die wesentlichen Erklärungspunkte sind wohl folgende:

1. Bei der Kleinheit der infrage kommenden Trägermengen kann nicht angenommen werden, daß sie phänomenal dicht sind (Forderung der Stetigkeit).
2. Wir haben keine Auskunft darüber, ob die Trägermengen strikt geordnete Ketten sind, d.h. ob die zum System $\mathcal{T} = \langle T,R \rangle$ gehörige Relation R ausschließlich Vor- bzw. Nach-Ordnungen zuläßt und nicht auch Gleich-Ordnungen (Forderung der Eindeutigkeit der Zuordnung).

An dieser Tatsache läßt sich auch nichts ändern, da es bei uns um die sozial-psychologische Gegebenheit der Bezugsgruppe (bzw. Mitgliedsgruppe) geht, die wir so nehmen müssen wie sie ist, und ohne daß wir vorher in irgendeiner Form wissen wie sie ist. Die Forderung an die Mitglieder der jeweiligen Gruppe, eine "Mitte" aufzusuchen, bezieht sich auf die "mittelste" Leistung des infrage stehenden Leistungskontinuums. Diese mittelste Leistung kann selbst aber nicht direkt befragt werden, sondern nur ein Element der Trägermenge, das hinsichtlich des betrachteten Merkmals-Kontinuums in der Mitte eines gegebenen Intervalls liegt. Bei nicht strikt geordneten Mengen wäre es dann möglich, daß mehrere Elemente (also hier: Mitglieder der Gruppe) am selben Ort oder/und sehr nahe beieinander liegen, sodaß bei Angaben über das mittelste Element eine ganze Reihe an Elementen genannt wird.

Das PFANZAGLsche Axiomen-System ist offensichtlich inadäquat für von uns betrachtete Mengen. Eine Abänderung wäre nötig derartig, daß es auch für finite Mengen verwndbar ist.

ii. Infinites-finites Axiomen-System (SUPPES & ZINNES).

Wie wäre eine passende Abänderung zu erreichen? Betrachten wir uns dazu folgende Axiomen-Systeme von SUPPES & ZINNES (28, S. 35ff):

a. Für den infiniten Fall:

"Definition ... A quaternary system $\mathcal{A} = \langle A, D \rangle$ is an infinite difference system if and only if the following seven axioms are satisfied for every a,b,c,d,e, and f in A:

1. IF abDcd and cdDef, then abDef,
2. abDcd or cdDab,
3. If abDcd, then acDbd
4. If abDcd, then dcDba, (+)
5. There is a c in A such that acDcb and cbDac,
6. If aPb and not abDcd, then there is an e in A such that aPe, ePb, and cdDae,
7. If aPb and abDcd, then there are e,f in A and an n such that $ceM^{n}fd$ and ceDab. "

(+) Es läßt sich zeigen, daß das Axiom 4 zu den Axiomen 1. und 3. nicht konsistent ist. Denn soll A4 gelten, dann gilt $baD^{-1}dc$ (D^{-1} gleich Inversion der Relation D). Soll die Relation D bedeuten "... nicht größer als ...", so wird sie entweder durch das A4 umdefiniert, oder es wird die Nicht-Kommutativität einer Distanz zwischen zwei Elementen hinsichtlich dieser Elemente gefordert, was sicher nicht beabsichtigt ist. Dann muß die Relation D eingeschränkt werden auf den Definitionsbereich D`gleich "... ist gleich groß wie...". (Forts. S.64)

b. Für den finiten Fall.

"Definition ... A finite equal difference system is a quaternary system $\mathcal{U} = \langle A, D \rangle$ in which A is a finite set, and for every a,b,c,d,e, and f in A the following five axioms are satisfied:

1. If abDcd and cdDef, then abDef;
2. abDcd or cdDab;
3. If abDcd, then acDbd;
4. If abDcd, then dcDba;
5. If aJb and cJd, then abDcd and cdDab.

In Axiom 5 the equal spacing assumption is imposed. "

> Zeichenerklärung:
> Die quarternäre Relation D bedeutet, daß die (subjektive) Differenz zwischen a und b kleiner oder gleich ist der Differenz zwischen c und d.
> Die Relation P (=preferred) aPb bedeutet, daß a einem b vorgezogen wird bzw. allgemeiner: vorgeordnet ist.
> aJb bedeutet, daß a der unmittelbare Vorgänger von b hinsichtlich der Relation P ist.
> Übrige Einzelheiten brauchen hier nicht zu interessieren.

In beiden Systemen sind die Axiome 1 - 4 dieselben. Das Axiom 5 im finiten Fall ist gegenüber dem infiniten Fall neu. Außerdem fallen gegenüber dem infiniten Fall die Axiome a5 - a7 (Existenz-Axiome) fort. Im PFANZAGLschen System müßte also das Axiom 1 ent-

---

Dann läßt sich allerdings A4 aus A3 ableiten und ist damit überflüssig:
Beweis:

Es sei	$abD'cd$	
dann	$acD'bd$	Axiom 3
dann	$bdD'^{-1}ac$	Inversion der Relation D'
dann	$baD'^{-1}dc$	Axiom 3
dann	$dcD'ba$	Inversion der Relation D'^{-1}
	qed.	

fallen bzw. ersetzt werden, was aber dann zur Folge hätte, daß das entscheidende Bisymmetrie-Axiom nicht mehr aufrechterhalten werden könnte. Die Axiome a5 - a7 bei SUPPES & ZINNES können aber nicht ersatzlos gestrichen werden, um auf den finiten Fall zu kommen. Statt dessen ist es nötig, da durch das Entfallen der Axiome a5 - a7 gewissermaßen metrische Information verloren geht, daß die Elemente der Trägermenge hinsichtlich des betrachteten Merkmals vorgeordnet werden.

Diese Forderung können wir aber nicht erfüllen, da es für uns, wie oben schon gesagt, ausgeschlossen ist, die Trägermenge hinsichtlich ihrer Merkmalsausprägungen zu manipulieren. Damit scheint es keine Möglichkeit zu geben, die Ergebnisse der Kategorisierungexperimente hinsichtlich der metrischen Voraussagen des WITTEschen Modells durch eine unabhängige Messung zu überprüfen, jedenfalls nicht durch eine fundamentale Messung (die zitierten Axiomen-Systeme von PFANZAGL und SUPPES & ZINNES sind beide solche für fundamentale Messungen).

(d) Skalierung durch abgeleitete Messung und Berechnung der Systemwerte.

Eine Überprüfung der Lage der metrisch relevanten Örter, der Bereichszäsuren, ist aber auf eine andere Weise möglich. Allerdings nicht mehr durch eine unabhängige Messung, sondern unter Verwendung der durch die Kategorisierung erhaltenen Daten. Nehmen wir nämlich die Lage der Zäsuren eines Systems

irgendeiner Höhe als gegeben hin unter der Annahme, daß sie an den geforderten Stellen des Intervalls [0,1] liegen, dann müssen die Zäsuren der Systeme von anderer Höhe an bestimmten Stellen liegen relativ zu jenen Zäsuren. Um diese relative Lage aber zu ermitteln, ist es nötig, die Distanzen zwischen den Beurteilungsgegenständen mindestens auf Intervallskalen-Niveau zu messen. Bei KELLNER führten derartige Überlegungen dazu, für die Abstandsbestimmung eine z-Skala zu wählen, da sich bei ihm die kategorialen Urteilshäufigkeiten mit hinreichender Annahme normal verteilten. Die sich pro Bereich ergebenden Anzahlen von Urteilsgegenständen wurden von Bereich zu Bereich kumuliert und auf die Gesamtanzahl auf das Intervall [0,1] normiert. Die so erhaltenen normierten Werte (der Zäsurstellen) wurden sodann einer z-Transformation unterzogen. Dieser Weg war für uns nicht gangbar, da eine Normalverteilung in den meisten Fällen nicht mit hinreichender Sicherheit angenommen werden konnte (s. Tab. Anhang V).

Da das Endziel der Skalierung in unserem Kontext eine empirische Überprüfung der THURSTONEschen Annahmen über die Lage der Zäsuren bei kategorialen Beurteilungen ist, diese Annahmen aber für den Fall einer linearen Verkettung der beurteilten Gegenstandsmengen aufgestellt wurden, nahmen wir eine lineare Skalierung vor.

Es schien zunächst nahezuliegen, die Berechnung der Skalenwerte der Zäsuren nach dem "law of categorical judgment" von THURSTONE vorzunehmen. Hierbei werden die kumulierten Prozentanteile der Urteilshäufigkeiten pro Kategorie und beurteiltem Gegenstand in z-Werte transformiert (unter einem z-Wert wird hier, wie üblich, der zu einem Prozentwert gehörige Abszissenwert unter der Verteilungsfunktion der Normalverteilung verstanden). Da die z-Werte, die den p-Werten von 0.0 und 1.0 entsprechen, unendlich groß werden - negativ bzw. posi-

tiv - , müssen sie in der erhaltenen Matrix der z-Werte (Matrix der Distanzen) zur weiteren Berechnung der Zäsuren gleich Null gesetzt werden. Je "trennschärfer" aber nun innerhalb eines Beurteilungssystems die Beurteilungen erfolgten, um so größer ist die Anzahl dieser mit Null besetzten Zellen der Distanz-Matrix. Die Anzahl dieser Null-Zellen (bzw. Leer-Zellen) überwiegt bei den vorliegenden Daten teilweise die der besetzten (einschließlich der Zellen mit p = 0.50, z = 0.00). Die Berechnung der Skalenwerte für die Zäsuren nach der Methode von DIEDERICHS et al. (s. SIXTL (26)) (+) für unvollständige Datenmatrizen (in unserem Falle unvollständige Distanzmatrizen) führte zu keinem Ergebnis. Offensichtlich wegen der zu hohen Zahl der mit Null besetzten Zellen der z-Matrix kam es zu keiner Konvergenz der Werte; von Iteration zu Iteration wurde das Iterationskriterium immer erheblich überschritten.

Die Zäsurenwerte lassen sich aber auch berechnen, wenn man die Abstände zwischen den beurteilten Gegenständen kennt und Aussagen über die Lage der Zäsuren innerhalb der Trägermenge machen kann. Wir berechneten die Skalenwerte der beurteilten Gegenstände (Personen) nach den Modellen von THURSTONE (in der Interpretation des "law of comparative judgment") und von BRADLEY-TERRY-LUCE (s. THURSTONE (30), TORGERSON (31), SIXTL (26)).

die Möglichkeit der Erstellung einer Häufigkeits-Dominanz-Matrix über die beurteilten Gegenstände aus kategorialen Daten darf

---

(+) Unter Verwendung eines ALGOL-Programms von SIXTL(26). Das Programm hatten wir für unsere Zwecke in FORTRAN IV übersetzt. An einigen Stellen waren Abänderungen nötig, da das Originalprogramm nicht ausreichend auf die Vermeidung von Null-Divisionen achtet.

als ohne Kommentar einsichtig angenommen werden. Ob allerdings, bei mathematischer Äquivalenz der Dominanz-Matrizen aus kategorialen und Paarvergleichs-Daten, auch eine psychologische Äquivalenz angenommen werden kann, bedarf eigener Untersuchungen.

Auch hierbei lassen sich allerdings Leerzellen in den Matrizen der Distanzen nicht vermeiden. Die Anzahl der Leerzellen nimmt hier ebenso zu mit zunehmender "Trennschärfe" der Beurteilungen. Da die jeweils mittlere Distanz zu allen anderen Gegenständen als Schätzwert für den Skalenwert eines jeweiligen Gegenstandes angenommen wird, ergibt sich somit eine verschiedene Sicherheit dieser Schätzung, wenn die Matrix der Distanzen Leerzellen enthält.

Für diese Schätzungen haben wir für beide Modelle jeweils vier verschiedene Methoden in der Behandlung der Distanz-Matrizen hinsichtlich der Leer-Zellen verwendet:

1. Die Leer-Zellen wurden nicht beachtet ("XQUER").
2. Die Leer-Zellen wurden beachtet derartig, daß die Summe der quadrierten Differenzen zwischen den Skalenwerten zweier Beurteilungsgegenstände und der Distanz zwischen diesen beiden ein Minimum wird, also (s. SIXTL):

$$S(i=1,N)\ S(j=1,N)\ (d(i,j)-R(i)+R(j))^2=\text{Minimum}$$

wobei: d(i,j) gleich Distanz zwischen den Reizen i,j (in z- oder log-Werten),

R(i),R(j) gleich Skalenwerte der Reize i,j (in z- oder log-Werten)

(i = 1,2,3,...,j,...,n)(+) ("MIN")

3. Das größte bzw. das kleinste Matrixelement wurde eingesetzt für den Fall von p=1.00 bzw. p=0.00. ("MAX")

(+) Für die Berechnung der Skalenwerte nach einem Lösungsverfahren von GULLIKSEN wurde eine FORTRAN IV-Version eines ALGOL-Programms von SIXTL verwendet.

4. Die Einsetzung wie im Fall3., aber mit dem Unterschied, daß entsprechend die Differenz zwischen größtem und nächst kleinerem Matrixelement hinzuaddiert bzw, diejenige zwischen kleinstem und nächstgrößerem subtrahiert wurde. ("MAX+D")

Die Einsetzungen 3. und 4. waren versucht worden, um zu überprüfen, inwieweit damit, neben dem Fall 2., eine Verbesserung der Skalierungsergebnisse erreicht werden kann. Sollten die Fälle 3. und 4. gegenüber dem Fall 2. wenigstens zu gleich guten Ergebnissen kommen, wäre ein erheblich ökonomischeres Verfahren zur Lösung des einschlägigen Problems gefunden.

Zum Vergleich der Güte der vier verschiedenen Einsetzungesmethoden wurde eine Konsistenz-Prüfung nach Mosteller vorgenommen.

Die Prüfung besteht im wesentlichen darin, daß aus dem Vektor der erhaltenen Skalenwerte die Matrix der relativen Urteilshäufigkeiten rückgerechnet wird und mit der Matrix der empirischen relativen Häufigkeiten verglichen wird. Errechnet wird ein Chi-Quadrat-Wert als Kenngröße über die durch eine Arcus-Sinus-Transformation in Bogengrade umgewandelten relativen Häufigkeiten beider Vergleichsmatrizen:

$$X^2 = \frac{S(Br(j,k) - Be(j,k))^2}{821/N}$$

wobei k größer j
Br(j,k) gleich Bogengrade des rückgerechneten Matrix-Elements E(j,k)
Be(j,k) gleich Bogengrade des empirischen Elements E(j,k)
N gleich Anzahl der Beobachtungen (d.h. in den meisten Fäälen gleich Anzahl der Versuchspersonen).

Klasse	Systemhöhe	z-Werte			log-Werte			
		MIN	MAX	MAX+1	XQUER	MIN	MAX	MAX+1
I	3	11.19	5.15	5.97	11.17	15.77	8.99	12.71
	4	7.64	3.79	4.05	5.94	8.94	5.04	6.85
	5	10.43	4.64	4.92	17.63	14.11	6.24	10.48
	6	9.25	5.41	4.83	12.39	10.93	5.40	7.21
II	3	7.27	3.67	4.08	6.65	10.48	6.29	8.49
	4	10.91	4.81	4.69	9.17	12.16	5.05	6.97
	5	11.70	4.87	4.59	11.94	12.75	4.95	7.34
	6	16.52	8.90	8.07	20.56	17.56	8.14	10.84
III	3	16.92	5.13	7.12	3033	20.08	8.55	16.72
	4	17.58	7.53	8.34	30.08	19.12	8.25	15.55
	5	20.24	11.59	11.12	28.10	19.54	9.85	14.38
	6	30.94	17.63	17.55	34.10	28.37	14.95	20.68
IV	3	13.19	4.23	5.10	25.77	18.74	7.17	13.22
	4	16.46	7.08	7.14	25.38	19.54	7.70	12.68
	5	14.62	7.25	6.27	40.00	17.43	5.81	10.33
	6	21.57	13.35	11.95	39.26	22.67	10.51	14.53
V	3	11.14	2.86	2.68	30.08	15.15	3.51	8.08
	4	14.27	7.53	5.78	61.88	17.71	5.35	8.44
	5	17.30	9.62	7.86	58.99	19.11	6.34	9.94
	6	19.26	15.75	12.72	93.50	20.68	10.93	12.37
VI	3	8.88	3.33	4.04	8.03	11.73	5.81	8.85
	4	8.11	3.88	4.12	5.29	8.81	4.48	6.16
	5	9.45	4.37	4.29	9.86	10.51	4.61	6.82
VII	3	13.63	4.13	4.21	41.74	18.32	5.17	12.68
	4	10.06	7.22	5.65	33.87	12.03	4.98	8.00
	5	16.51	10.62	9.12	35.08	15.68	7.06	10.52
Spaltenmittel		14.04	7.09	6.78	27.91	16.09	6.96	10.80

Tab. 7. Mosteller-Koeffizienten.

MOSTELLER selbst scheint diese Kenngröße zur Linearitäts-Prüfung entwickelt zu haben (s. SIXTL). Aus verschiedenen Gründen, die hier nicht diskutiert zu werden brauchen, wird sie hiermit jedoch überfordert. Nur für den Fall, daß die Kenngröße gleich Null wird, läßt sich strikt auf Linearität schließen. Möglich erscheinen aber komparative Aussagen über die Güte verschiedener Meßmodelle (in Anwendung über dieselbe Trägermenge einer Eigenschaft). Tab. 7 zeigt die Kenngrößen dieser Prüfung für die verschiedenen Klassen und Systemhöhen bezüglich beider Modelle und der vier Ersetzungsmethoden. Die Schätzung der Skalenwerte bei einfacher Mittelwertsbildung ohne Beachtung der Leerzellen ist deutlich schlechter als die der anderen drei. Die besten Schätzungen lassen sich durch die Methoden 3. und 4. erzielen. Die Unterschiede zwischen diesen beiden Methoden und die Unterschiede in diesen beiden Methoden zwischen den beiden Modellen lassen sich wohl nicht interpretieren, da die geringfügigen und in der Richtung teilweise auch uneinheitlichen Differenzen möglicherweise zurückzuführen sind auf die unterschiedliche Behandlung der für die Hin- und Rück-Transformation der relativen Häufigkeiten verwendeten Funktionen. D.h. Argument der einen Funktion und Wert der zugehörigen Umkehrfunktion sind nicht identisch, also:

$$a = F(b)$$

und

$$b = F_U(a)$$

führt nicht zu

$$b = b \quad .$$

Die Darstellungen der Funktionen sind uns nicht bekannt. Es handelt sich um die "Bibliotheks-Funktionen" ERF(x), ALOG(x), EXP(x) der benutzten elektronischen Rechenanlage. Nach unseren Erfahrungen ergeben sich stärker differierende Ergebnisse bei extremen Werten.
Im übrigen ist zu erwähnen, daß die Skalenwerte über dieselbe Trägermenge nach beiden Modellen sich nahezu linear aufeinander abbilden mit einem Steigungskoeffizienten kleiner 1 bei Abbildung der Modelle BTL auf THURSTONE.

Die Kenngrößen können benutzt werden für die Auswahl derjenigen Ersetzungsmethode, aufgrund deren Skalenwerte die Systemwerte berechnet werden. Diese haben wir allerdings zum Vergleich nach allen vier Methoden berechnet. Aus den Abbildungen 5 - 9, 10,11 läßt sich die jeweilige Lage (topologischer Art) einer Zäsur ablesen. Die Zäsurwerte wurden so berechnet, daß der entsprechende Skalenwert des jeweiligen Elements der Trägermenge, auf dem Zäsur lag gewählt wurde. Lag die Zäsur zwischen zwei Elementen, dann wurden die Skalenwerte dieser beiden einfach gemittelt. Die Systemwerte für die Zäsuren wurden wie bei KELLNER berechnet: und zwar für die Distanz von der vermuteten System-Mitte her, die Zäsurwerte der Dreier-Katetorisierung dienten als Bezugswerte:

$$S(x) = \frac{E(x)\ S(x)}{E(3)}$$

wobei $S(x)$ gleich empirischer Systemwert der fraglichen Zäsur;

$S(3)$ gleich zu $S(x)$ "vorzeichengleicher" theoretischer Systemwert der Dreier-Kategorisierung;

$E(x), E(3)$ gleich Skalenwerte der entsprechenden Zäsuren.

Die Tabellen 1 - 5, 7, 11 zeigen die empirischen Skalenwerte und die gefundenen Systemwerte für diejenigen Zäsuren, deren Lage aus den zugehörigen Abbildungen erkennbar waren.

7. Befunddiskussion.

Die Wahl der theoretischen Zäsurwerte der Dreier-Kategorisierung als Bezugswerte für die Berechnung der übrigen Zäsurwerte ist beliebig. Die Zäsurwerte jeder andersstufigen Kategorisierung hätten auch dazu dienen können. Zum Vergleich der empirischen Systemwerte mit den aus ~~aus~~ dem WITTEschen Modell geforderten theoretischen standen also nur die Werte der Zäsuren der Vierer-, Fünfer- und Sechser-Kategorisierung zur Verfügung.

Bei diesem Vergleich zeigen sich zwar von Fall zu Fall verschieden große Differenzen. Eine gewisse Affinität der empirischen zu den theoretischen Werten ist aber unverkennbar. Dieser Sachverhalt ist um so erstaunlicher, wenn man die Menge der Umformungen der Ausgangsdaten bedenkt, die zur Berechnung der Werte notwendig war. Darüberhinaus waren die vorgenommenen Schätzungen oft recht grob (z.B. einfache Mittelung der Skalenwerte, wenn eine Zäsur zwischen zwei Beurteilungsgegenstände zu liegen kam). Zum anderen waren die Modellvoraussetzungen (jedenfalls für die Skalenberechnung nach dem Modell von THURSTONE) empirisch nicht überprüft worden. Für die Verwendung des Falls V des THURSTONEschen Modells kann man sich freilich darauf berufen, daß (a) Untersuchungen gezeigt haben (s. SJÖBERG (27)), daß die Skalen nach dem FALl IV und dem Fall V über dieselbe Trägermenge sich, zumindest im mittleren Bereich, nahezu linear abbilden und die Datenkonsistenz nach dem Fall IV nur gerinfügig verbessert wird gegenüber dem Fall V. Bedeutsamer ist aber, daß (b) nach

MOSTELLER (s. TORGERSON (31), S. 179), wenn von n Reizen sich bei k Reizen die Urteile im Sinne der Modellvoraussetzungen des Falls V verteilen (d.i. sich normal verteilen), n-k Reize aber nicht, diese k Reize relativ zueinander modellentsprechend richtig distanziert sind, nicht aber zu den übrigen n-k Reizen, diese auch nicht untereinander. Wir können aber annehmbar behaupten, daß im Falle der Zäsuren die für ihre Berechnung benötigten Reize hinsichtlich ihrer Urteilsverteilungen in etwa die Modellvoraussetzungen erfüllen, die anderen Reize dagegen nicht. Somit sind für die hier allein interessierenden Zäsuren keine oder keine sehr groben Verstöße gegen die Modellvoraussetzungen anzunehmen.

Wir verzichteten von vornherein auf die Berechnung der Skalenwerte nach dem adäquateren Fall IV des THURSTONEschen Modells, da der Gewinn (wenn überhaupt) nur geringfügig gewesen wäre, der Rechenaufwand aber unverhältnismäßig viel größer als er es sowieso schon war.

(a) Zufallskritische Überprüfung der Differenzen zwischen empirischen und theoretischen Systemwerten.

Eine Berechnung von Konfidenzintervallen, mit den üblichen Annahmen über normalverteilte Fehlerstreuung von Meßwerten, zur Prüfung der Relevanz der gefundenen Differenzen zwischen empirischen und theoretischen Zäsurlagen erscheint nicht sinnvoll, da bei einer Kombination der Intervallgrenzen von der zu prüfenden mit der "Bezugs"-Zäsur das sich ergebende Intervall um die zu prüfende

Zäsur immer einen erheblichen Bruchteil (im Durchschnitt um 10%) des gesamten Systemumfanges ausmachen würde. Bei einer derartig großen Toleranz läßt sich keine vernünftige Entscheidung über die Frage nach der Modellübereinstimmung der gefundenen Ergebnisse treffen.

KELLNER verwendete eine Berechnung zur Prüfung der Abweichung der empirischen von den theoretischen Zäsurlagen ("theoretische Zäsurlage" soll hier natürlich immer heißen: vom WITTEschen Modell her geforderte Zäsurlage), die zwar noch einer allgemeinen Begründung bedarf, aber zunächst eine gewisse Plausibilität besitzt. Nimmt man an, daß der durchschnittliche Abstand benachbarter Elemente einer Trägermenge auf dem jeweils betrachteten Merkmalskontinuum 1/(n-1)tel des Systemumfangs beträgt (wobei n gleich Anzahl der Elemente), dann kann man diesen Abstand als einen Bereich betrachten, innerhalb dessen eine Abweichung von den theoretischen Werten zulässig ist. Für unsere Fälle hieße das, daß Abweichungen zwischen 2.4 und 3.3 in "Systemeinheiten" zulässig sind. Wie aus den Tabellen 1 - 5, 8, 12 ersichtlich, wird dieser Toleranzbereich nicht immer eingehalten.

Eine derartige Berechnung eines Zulässigkeitsbereichs (der im übrigen nicht statistischer Art ist) gerät aber in Schwierigkeiten, wenn nicht die begründete Annahme gemacht werden kann, daß die zugrunde liegende Verteilung der Elemente der Trägermenge über das betrachtete Merkmalskontinuum nur zufällig von einer Gleichverteilung abweicht. Die Berechnung kann aber sinnvollerweise nicht vorgenommen werden bei gleichzeitiger Annahme (und sogar Behauptung) einer "Normalverteilung"

der Elemente über das Merkmalskontinuum.

Irgendwelche Annahmen über die Anordnung der Elemente auf dem Merkmalskontinuum führen darüberhinaus auch nicht weiter. Denn entscheidend für die Bestimmung eines Zulässigkeitsbereichs ist nicht die Abständigkeit und Dichte der Trägermenge, also Annahmen über das Gemessene, sondern Annahmen über die Genauigkeit des Meßinstruments. Dieses sind hier die kategorialen Urteile; die Genauigkeit des Meßinstruments hängt von der Anzahl der Urteile ab. Diese Aussage ist zunächst nicht im statistischen Sinne zu verstehen, etwa daß die Genauigkeit der Schätzung eines Verteilungsparameters von der Größe der Stichprobe abhängt, sondern im Sinne der Frage, bei wie kleinen Differenzen im Gegenstandsbereich isomorphe Abbildung der Strukturen im Zahlenbereich möglich ist. Je genauer das Meßinstrument, desto kleinere Differenzen im Gegenstandsbereich werden erfaßt. Das Maximum an Genauigkeit im Falle von Beurteilungen wird dann bestimmt durch den Grad der Veränderung eines Meßwertes bei einer gegebenen Meßfunktion durch die Veränderung eines Urteils. Da beide hier verwendeten Funktionen (Verteilungsfunktion der Normalverteilung, logistische Funktion) normierte Urteilsmengen messen, ist es klar, daß bei einer Vergrößerung der Menge der Urteile (entweder durch eine Vergrößerung der Anzahl der Beurteiler oder der Urteilswiederholungen oder beides) die Meßgenauigkeit erhöht wird. Bei den bekannten Eigenschaften beider Funktionen wirkt sich bei einer gegebenen Urteilsmenge die Veränderung eines Urteils verschieden aus und zwar so, daß zu den Rändern der Funktionen hin die Meßwerte stärker zunehmend verändert werden.

Z.B.: nehmen wir der Einfachheit halber an, die Veränderung eines Urteils über einen beurteilten Gegenstand mache bei Normierung der Urteilsmenge auf das Intervall $[0,1]$ $p = 0.01$ aus. Dann würde sich im Falle der Normalverteilung als verwendeter Meßfunktion im mittleren Bereich bei einer Veränderung von $p = 0.50$ auf 0.51 der entsprechende Meßwert von $z = 0.000$ auf $z = 0.025$, im Randbereich bei einer Veränderung von $p = 0.97$ auf $p = 0.98$ der Meßwert von $z = 1.881$ auf $z = 2.054$ verändern. D.h. im mittleren Bereich tritt eine Differenz von 0.025 auf, im Randbereich aber eine von 0.173 i.E. der Streuung auf. Die Funktion mißt also an ihren Rändern weniger genau als in der Mitte. Das bedeutet für unsere Fragestellung, daß je exzentrischer die Zäsuren liegen, desto größer die Zulässigkeitsbereiche werden.

Werden nun für die Berechnung der Systemwerte der untere Grenzwert des Zulässigkeitsbereichs der Kriteriums-Zäsur mit dem der Vergleichs-Zäsur und die beiden jeweiligen oberen Grenzwerte miteinander kombiniert, so ergeben sich auch hierbei derartig große Zulässigkeitsbereiche für die Lage der betrachteten Zäsuren, daß gewissermaßen "alles" damit bewie sen werden kann. D.h. die durch das WITTEsche Modell geforderten Zäsuren-Lagen lassen sich auf diese Weise belegen.

(b) Normalverteilung kategorialer Urteile.

Die Überprüfung der Lage der Zäsuren über die empirischen Systemwerte führt also bei unseren Daten zu keinem Erfolg sowohl
(1) durch Vergleich der Ergebnisse des Kategorisierungsexperiments mit einer fundamentalen Messung (PFANZAGL-Versuch) als auch
(2) durch abgeleitete Messung, da die verwendeten Meßfunktionen für unsere Zwecke zu ungenau messen.

Allein auf dem Niveau minimaler Forderungen an eine Skala, auf topologischem Niveau, läßt sich (freilich auch nicht in allen Fällen) isomorphe Abbildung erreichen in den Fällen, in denen eine Bereichszäsur überhaupt ablesbar war.

Es läßt sich nun fragen, ob sich aus dem WITTEschen Modell in Verbindung mit den bei unseren Daten vorliegenden Bedingungen Annahmen herleiten lassen, von denen her eine Modellkonformität der Daten sich aufweisen läßt, auch hier natürlich unter der Voraussetzung, daß im jeweils zu diskutierenden Fall bei gegebener Systemhöhe ausreichende Bereichsdifferenzierung stattgefunden hat. Nehmen nach dem WITTEschen Modell auf dem phänomenalen Merkmalskontinuum die kategorialen Bereiche gleich große Abschnitte ein, so heißt das, daß bei phänomenal äquidistanter Reizanordnung die (kategorialen) Urteilsmengen gleich mächtig sind. Ein Bezugsystem besitzt also Symmetrieeigenschaften; entsprechend sind die Urteilsmengen symmetrisch verteilt.
Wenn man von Leistungen annimmt, daß sie sich normal verteilen, dann müssen die (kategorialen)

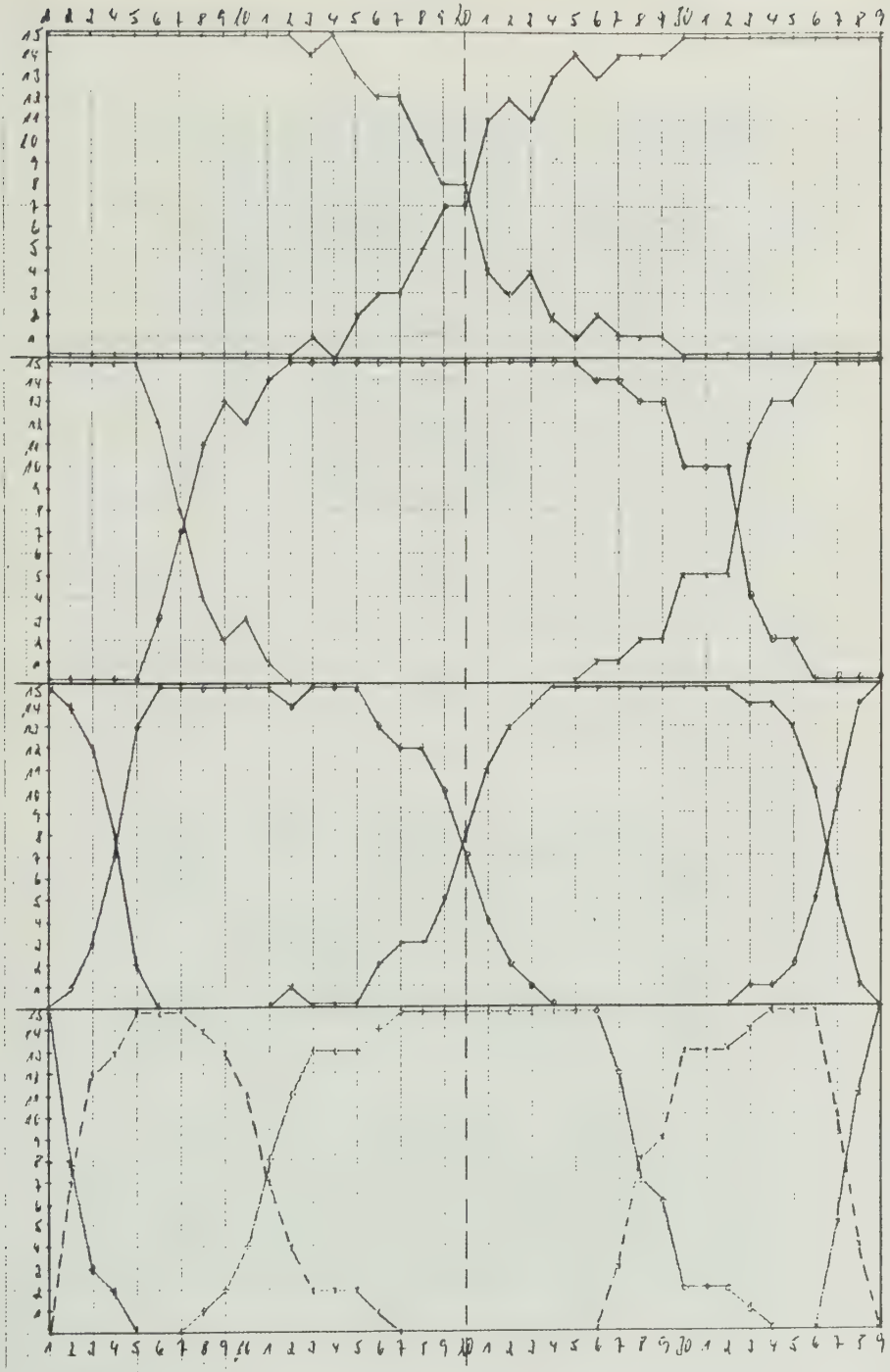


Abb. 10

Urteilsmengen bei Beurteilungen von Leistungen von mittleren zu Rand-Kategorien abnehmen. Abb. 10 zeigt das Ergebnis eines Modellexperiments mit konstruierter "Normalverteilung" der Elemente.(Über Details des Experiments s. Anhang IV.) Die kategorialen Bereiche auf dem betrachteten Merkmalskontinuum sind gleich groß, so daß die Urteilsmengen von den mittleren zu den Rand-Bereichen hin abnehmen wie erwartet. (In der Abbildung sind die Elemente der konstruierten Trägermenge entsprechend ihrer Größe in eine Rangreihe gebracht ohne Berücksichtigung der sonstigen metrischen Verhältnisse.)

(c) Zum Begriff des "mnemisch stabilisierten Systems".

Bevor wir aber dieses Argument der Normalverteilung weiter verfolgen, läßt sich ein anderes zur möglichen Erklärung der relativen Erfolglosigkeit der bisherigen Verifizierungsversuche mit unseren Daten anbringen. Man kann annehmen, daß es sich bei den untersuchten Systemen im Sprachgebrauch WITTEs um "stabilisierte" Systeme, genauer um "mnemisch stabilisierte" Systeme handelt.

Was bedeutet nun dieser Begriff? Wir werden den Versuch einer vollen Explikation dieses Begriffs machen, ohne behaupten zu wollen, daß alle Momente bei WITTE auch implizit mitgemeint waren. Nehmen wir die obige Darstellung eines Bezugsystems $\mathfrak{S}$ als eines geordneten Tripels $\mathfrak{S} = \langle \mathcal{T}, \mathcal{E}, Z \rangle$

bestehend aus den Elementen $\mathcal{T}$ als System von über ein Merkmal geordneten Gegenständen, aus $\mathcal{E}$ als System von zu diesem Merkmal konformen geordneten Eigenschaften und einer Zuordnung Z der Eigenschaften zu den Gegenständen, so bestehen grundsätzlich zwei Möglichkeiten einer Stabilisierung: 1. auf der Gegenstandsseite, 2. auf der Beurteilerseite (denn es handelt sich hier zunächst immer um "Beurteilungen", zu denen auch ein Beurteiler gehört; die Frage nach nicht verbal gebundenen Eigenschaftensystemen oder tierischen Systemen braucht jetzt nicht diskutiert zu werden). Für die 1. Möglichkeit läßt sich nur schwer für eine gegebene Fragestellung im allgemeinen eine Antwort geben; sie gehört zur Ökologie der Bezugsysteme. Man kann aber die Annahme von relativ konstanten alltäglichen Umweltbedingungen machen. Diese relative Konstanz der alltäglichen Umweltbedingungen ist die Voraussetzung für die 2. Möglichkeit, d.h. dafür, daß sich "mnemisch stabilisierte" Systeme aufbauen können.

Doch was stabilisiert sich mnemisch? Im allgemeinen wird angenommen, daß ein in irgendeiner Hinsicht minimales und maximales Element einer geordneten Menge sich mnemisch sedimentiert, und daß die Stabilität dieses mnemischen Sediments die Stabilität des Bezugsystems ausmacht. Jene beiden Elemente wirken sich konstitutiv für die Struktur eines Bezugsystems aus. Das wäre eine Möglichkeit (a). Unter Berücksichtigung des angegebenen formalen Begriffs des Bezugsystems besteht eine zweite Interpretationsmöglichkeit (b): das mnemische Sediment enthält nicht nur die beiden Randelemente, sondern die Trägermenge in ihrer Dichte, Verteilung und/oder sonstigen Kenngrößen, d.h. es enthält auch die beiden Randelemente. D.h. die Strukturierung

eines Bezugsystems konstituiert sich zwar von den Randelementen her, aber unter "Berücksichtigung" der Verteilung der Elemente der Trägermenge über ein betrachtetes Merkmalskontinuum (das ist der Ansatz PARDUCCIs). Weiterhin (c): es besteht nicht nur ein mnemisches Sediment von "Kenngrößen" einer Trägermenge, sondern auch hinsichtlich einer Zuordnungsmöglichkeit ein Sediment eines Eigenschaftensystems und zwar in Form eines bestimmten Differenzierungsgrades bei einer spezifischen Verbalisierung der Elemente dieses Systems, also der Eigenschaften (oder Kategorien).

"Mnemisch stabilisiertes System" kann also heißen: stabilisierte Zuordnung eines spezifizierten Eigenschaften-Systems zu einem (meistens aus längerer Erfahrung gegebenen) Gegenstands-Systems. Liegt also in einem gegebenen Fall ein derartig stabilisiertes System vor, so besteht keine Beliebigkeit der experimentell vorgegebenen (a) zu beurteilenden Trägermenge, (b1) Mächtigkeit der Menge der Eigenschaften des Eigenschaften-Systems, (b2) der verbalen Spezifikation dieser Eigenschaften. In Abb. 11 wird die Auswirkung einer von den übrigen Erhebungen abweichenden verbalen Spezifikation der Eigenschaften einer fünfstufigen Kategorisierung gezeigt. Anstelle der sonstigen Eigenschaften: sehr gut, gut, mittel, schlecht, sehr schlecht - wurden folgende verwendet: gut, eher gut, mittel, eher schlecht, schlecht.
Zur vierstufigen Kategorisierung hin ergibt sich keine Verschiebung der "gut"- und "schlecht"- Bereiche, so daß zu vermuten ist, daß diese beiden Randbereiche sich als jene stabilisiert haben, genauer, daß die Bereichsgrenzen der beiden Randbereiche

zu den benachbarten inneren Bereichen sich stabilisiert haben, so daß hier die Zuordnung der Eigenschaften außerordentlich wenig trennscharf erfolgt (zu beachten ist allerdings die dabei auffällig zu nennende Stabilität der zu vermutenden System-Mitte). Die beiden möglicherweise für eine trennscharfe fünfstufige Kategorisierung erforderlichen Extremkategorien "sehr gut","sehr schlecht" werden natürlich von den bestehenden Randbereichen absorbiert.

Die Möglichkeit der Stabilisierung eines Systems auf einem n-stufigen Differenzierungsgrad bei spezifischer Verbalisierung der Eigenschaften ist dort zu vermuten, wo, wie in unserem Falle, Beurteiler und beurteilte "Gegenstände" dieselben sind, das Merkmal Relevanz besitzt,und die Art der Beurteilung verhaltenskonsequent ist, oder auch wenn die Beurteilung "verhaltensfunktional" ist. D.h. wenn die Art der Beurteilung bedingt, wer, wie bei "Güte im Fußballspiel", zur Klassenmannschaft gehören soll und wer nicht, oder daß derjenige, der zur Mannschaft gehört, zu den "Guten" gehört, der andere aber zu den "Schlechten". Im letzteren Fall ist also eine Beurteilung gruppendynamisch bedingt. Im Extremfall kann das bedeuten, daß das Eigenschaften-System nur alternativ ist (das in Abb. 7 gezeigte System scheint solch ein Fall zu sein).

Ist ein stabilisiertes Bezugsystem $\mathfrak{L} = \langle \mathcal{T}, \mathcal{E}, z \rangle$ bezüglich des Eigenschaften-Systems singulär (das soll hier als Begriff eingeführt werden) und erfüllen die hierbei entstehenden metrisch relevanten Örter (Pole, Zäsuren) metrische Bedingungen (irgendeines metrischen Axiomen-Systems),

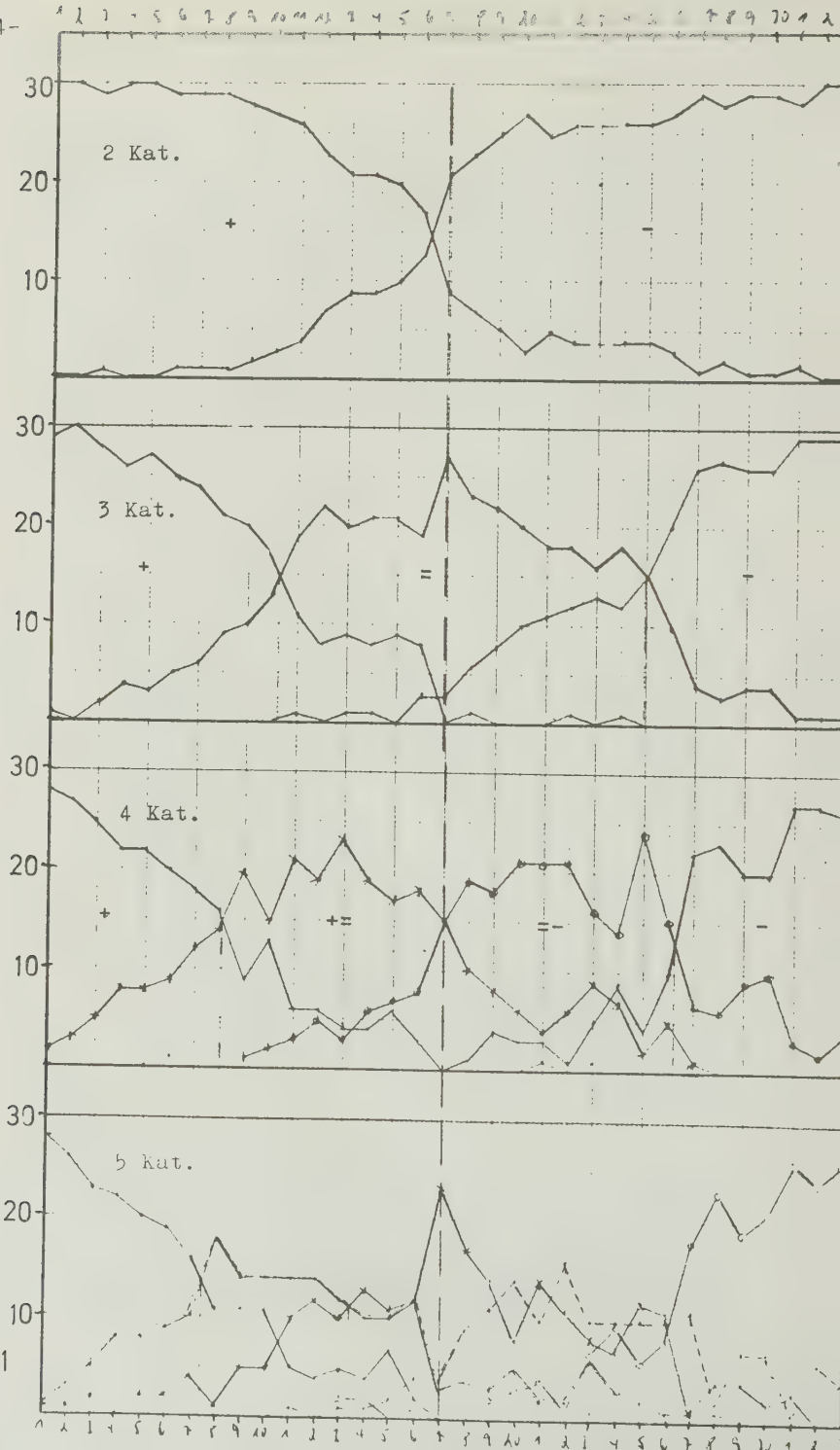


Abb.11

Klasse VI		Z-Werte				log-Werte			
		MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Lage d. Zäsuren nach d.empir.Skalenwerten									
Zäsur:									
1/2		1.243	-0.108	-0.076	-0.073	2.279	-0.184	-0.122	-0.111
1/3		0.633	-0.622	-0.715	-0.729	1.166	-1.200	-1.311	-1.356
2/3		2.033	0.547	0.743	0.768	3.679	0.975	1.359	1.448
1/4		0.391	-0.724	-0.942	-0.973	0.753	-1.288	-1.718	-1.829
2/4		1.431	0.063	0.127	0.133	2.622	0.129	0.252	0.274
3/4		2.310	0.813	0.9922	1.015	4.181	1.455	1.810	1.911
Systemwerte:									
Zäsur:	erwartet:								
1/4	-25.0	-21.7	-19.1	-21.1	-21.4	-21.4	-18.0	-21.0	-21.5
3/4	25.0	24.1	25.9	23.2	23.0	25.5	26.1	23.4	23.2

Tab. 9. Zäsurenwerte.

dann kann nicht erwartet werden, daß Systeme über dasselbe Merkmalskontinuum und dieselbe Trägermenge aber von andersstufigem Differenzierungsgrad, der je experimentell eingeführt wird, dieselben metrischen Bedingungen erfüllen selbst dann nicht, wenn das System ausreichend trennscharf gegliedert erscheint. Für sogenannte rezente Systeme braucht dieser Sachverhalt nicht zuzutreffen.

Handelt es sich bei uns in allen Fällen um derartige bezüglich des Eigenschaften-Systems singuläre Bezugsysteme, dann haben alle berechneten Systemwerte für die Verifizierung oder Falsifizierung des WITTEschen Modells keine Beweiskraft gleichgültig, auf welche Weise die empirischen Zäsurwerte berechnet wurden und ob sie mit Meßfehlern irgendwelcher Art belastet sind oder nicht.

Allein aus den Ergebnissen von Kategorisierungsexperimenten läßt sich allerdings nicht entscheiden, welches System im gegebenen Fall das "eigentliche" Bezugsystem darstellt. Hierzu bedarf es weiterer metrischer Information unabhängig vom Kategorisierungsexperiment vorausgesetzt, daß man nur das formale Kriterium der Übereinstimmung mit einem metrischen System-Modell (wie das WITTEsche eines ist) heranzieht.

(d) Verteilungsform kategorialer Urteile.

Es läßt sich nun allerdings fragen, ob überhaupt eines der dargestellten Systeme Bedingungen eines metrischen System-Modells erfüllt. Das führt zurück zur obigen Diskussion um eine "Normalverteilung" der kategorialen Urteile. Kann überhaupt Normalverteilung erwartet werden? Normalverteilung kann angenommen werden von Grundgesamtheiten (oder sie - im statistischen Sinne - in ausreichendem Maße repräsentierenden Stichproben). Was wäre in unserem Falle die Grundgesamtheit? ist der statistische Begriff der Grundgesamtheit überhaupt vereinbar und/oder relevant für den Begriff des Bezugsystems (wir setzen hier natürlich voraus, daß für empirische Verteilungen die Normalverteilung sowieso nur annäherungsweise Gültigkeitscharakter besitzt, d.h. im gegebenen Fall für praktisch-empirische Zwecke in ausreichendem Maße eine empirische Verteilung beschreibt)? Für den statistischen Begriff der Gesamtheit ist die Zusammenfassung der Elemente zu dieser Gesamtheit beliebig. Es braucht hier nicht diskutiert zu werden, daß im gegebenen Fall die Mengen (Trägermengen) immer irgendwie sinnvoll zusammengefaßt werden. Grundsätzlich ist aber die Zusammenfassung bis auf ein betrachtetes Merkmal beliebig.

Zur Annahme einer Verteilungsform reicht es aus, daß alle Elemente der Grundgesamtheit ein gemeinsames Merkmal besitzen, dieses Merkmal (stetig) abstufbar ist, und jedes Element durch einen Abstufungsgrad dieses Merkmals beschrieben werden kann (jedenfalls bei eindimensionalen Verteilungen). In einem Bezugsystem $\mathfrak{B} = \langle \mathfrak{T}, \mathfrak{E}, Z \rangle$ ist aber die Menge

der Gegenstände des Gegenstands-Systems $\mathcal{T}$ nicht beliebig. Zusammengefaßt werden nur die Elemente, die als zu einem Bezugsystem zugehörig erlebt werden. Der Unterschied besteht also in der Nicht-beliebigkeit der Zusammenfassung. Dieser Sachverhalt beinhaltet, daß die Forderung der "großen Zahl" gar nicht erfüllbar ist, obzwar im gegebenen Fall in ausreichendem Maße erfüllt sein kann. Damit besteht keine Möglichkeit, a priori irgendwelche Verteilungen einer Grundgesamtheit anzunehmen. Die Art der Verteilung der Elemente der zu einem Gegenstands-System gehörigen Trägermenge über ein betrachtetes Merkmalskontinuum kann immer nur post hoc festgestellt werden.

Ist die erlebte Ähnlichkeit der Elemente der Trägermenge T über ein Merkmalskontinuum, d.h. die phänomenale Distanz, invariant gegenüber der Verteilung der Menge über das Merkmalskontinuum, dann kann die (normal- oder gleichverteilte) Symmetrie der Zuordnung Z der Elemente E des Eigenschaften-Systems $\mathcal{E}$ zu den Elementen der Trägermenge T des Gegenstands-Systems $\mathcal{T}$ eines Bezugsystems $\mathcal{B}$ allein kein Kriterium dafür sein, daß die metrische Struktur eines metrischen System-Modells empirisch gegeben ist.

(e) Invarianz von Distanzen bei variierendem Verteilungstyp (PARDUCCI).

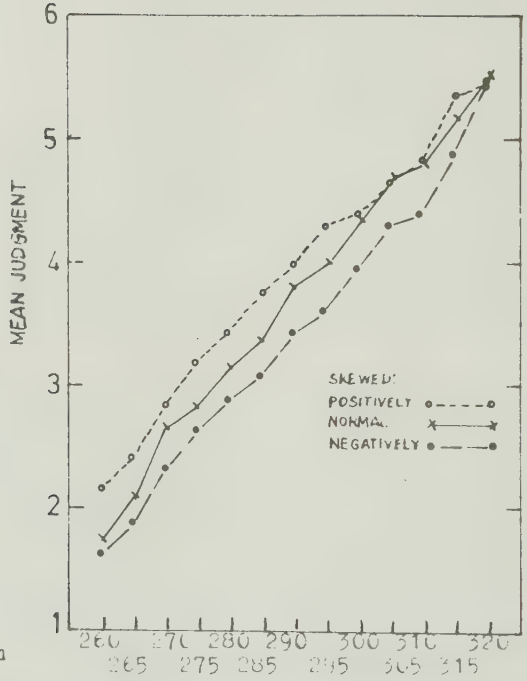
Ob allerdings die erlebte Ähnlichkeit - oder phänomenale Distanz - der Elemente der Trägermenge eines zu einem Bezugsystem gehörenden Gegenstands-Systems bei gegebenem Merkmalskontinuum invariant

gegenüber der Art der Verteilung dieser Elemente ist oder nicht, ist noch kontrovers..

Nach PARDUCCI (19,20) geschieht die Zuordnung der Eigenschaften eines Eigenschaften-Systems zu den Elementen der Trägermenge derartig, daß bei Bestehen zweier "Tendenzen" diese jeweils einen Kompromiß eingehen, einen "frequency-range compromise". D.h. er nimmt an, daß der zwischen einem irgendwie minimalen und maximalen Element der Trägermenge bestehende Sektor eines Merkmalskontinuums durch Zuordnung von Eigenschaften in gleich große Abschnitte zerlegt wird. Andererseits wird aber auch die Trägermenge durch die Zuordnung in gleich große Teilmengen zerlegt. Ist die Verteilung der Elemente der Trägermenge auf dem Merkmalskontinuum aber derartig, daß eine Zerlegung des betrachteten Sektors in gleich große Abschnitte die Trägermenge in ungleich mächtige Teilmengen zerlegen würde, so gehen beide Tendenzen einen "Kompromiß" ein.

> Hierzu s. Abb. 12. Der Verlauf der Kurven zeigt, daß die durchschnittliche Zuordnung abhängig ist von der Art der Verteilung. Bei linksschiefen Verteilungen werden leise Geräusche als lauter beurteilt als bei normalen oder rechtsschiefen Verteilungen - natürlich bei entsprechender Anordnung der Reize.
> Wie dieser Kompromiß im einzelnen aussieht, ist, soweit wir übersehen, von PARDUCCI nicht dargestellt worden. Es besteht die Möglichkeit, daß er ein Minimumsproblem darstellt, d.h. daß die Summe der Abweichungen von durch beide Tendenzen erwarteten Werten ein Minimum darstellt, wobei eine verschiedene Gewichtung der beiden Tendenzen nicht ausgeschlossen ist.

Die Behauptung PARDUCCIs vom "frequency- range compromise" besitzt möglicherweise ihren Geltungsbereich nur bei schiefen Verteilungen. Das Egebnis



dB-sum

Frequency:

pos.skew.	15	13	13	12	11	10	8	6	5	4	3	3	3
normal	1	2	4	8	12	16	17	16	12	8	4	2	1
neg.skew.	3	3	3	4	5	6	8	10	11	12	13	13	13

Mean judgments of stimulus sets in the three distributions ... Each set is characterised by the sum of its four dB components as listed on the abscissa.

Abb. 12. Aus HANDEL et al. (21, S. 146).

unseres Experiments, das wir in Abb. 10 dargestellt haben, zeigt deutlich, daß (jedenfalls bei symmetrischen Verteilungen) kein derartiger Kompromiß stattfindet. Die kategorialen Bereichsabschnitte sind auf dem betrachteten Kontinuum gleich groß, während die durch die Abschnitte gebildeten Teilmengen der herangezogenen Trägermenge deutlich ungleich mächtig sind. Hat aber PARDUCCIs Modell Gültigkeit für den Fall schiefer Verteilungen, d.h. geschieht in einem solchen Fall die Zuordnung der Eigenschaften zu den Gegenständen nicht ausschließlich nach dem phänomenalen Ort, den die Gegenstände auf dem betrachteten Merkmalskontinuum einnehmen, so heißt das, daß diese Zuordnung keiner metrischen Interpretation zugänglich ist.

Da bei mehreren der von uns untersuchten Gruppen derartige schiefe Verteilungen vorzuliegen scheinen, ist es von den Überlegungen PARDUCCIs ausgehend nicht zu erwarten, daß die Bedingungen metrischer System-Modelle erfüllt werden können. Eine metrische Interpretation wäre allerdings dann möglich, wenn man davon ausgeht, daß bei schiefen Verteilungen, also bei einer exzentrischen Verlagerung des Zentrums der Verteilung, gleichzeitig die Ähnlichkeitsverhältnisse der Gegenstände untereinander mitverändert würden. Nur bliebe die Schwierigkeit zu erklären, warum im Falle einer "Normal"-Verteilung, also einer symmetrischen Verteilung, derartige Veränderungen nicht vorkommen. -
Hier besteht noch ein weites Feld für weitere Untersuchungen.

(f) Invarianz von Distanzen bei variierender Position der Beurteiler.

Die bisher dargestellten Annahmen gehen davon aus, und das ist auch Annahme der"klassischen" Meßtheorie, daß,von Fehlervarianz abgesehen - mit eingeschlossen die Varianz zwischen messenden Systemen (wie etwa beobachtenden, beurteilenden Personen), der Unterschied zwischen den Meßwerten auf Unterschiede zwischen den gemessenen Objekten zurückzuführen sind. Unterschiede zwischen den Meßwerten jeweils eines Objekts , bei Berücksichtigung der Tatsache, daß die Meßwerte - außer bei topologischen Skalen - immer nur eindeutig sind bis auf lineare Transformation, sind zurückzuführen auf Störgrößen, Inadäquatheit der Meßapparate und/oder der Meßmethoden entweder aller oder einiger im Vergleich mit anderen. Selbst bei praktischer Verunmöglichung einer Messung bleibt die Vorstellung, daß Unterschiede zwischen den Objekten eindeutig abbildbar sind. Die Unterschiede zwischen den Objekten bestehen für sich, der Vorgang der Abbildung verändert sie nicht, solange diese "richtig" ist.

Diese Vorstellung kann aber nicht aufrecht erhalten werden. Die Ergebnisse der "social-perception"-Forschung haben häufig gezeigt, daß systematische, d.h. voraussagbare Veränderungen in der wahrnehmungsmäßigen und/oder kognitiven Erfassung von Objekten stattfinden unter veränderten Zusatzbedingungen meist wertmäßiger Art. Derartige Veränderungen ließen sich auch bei Beurteilungen von Gegenständen hinsichtlich eines Merkmalskontinuums erreichen.

TAJFEL (29) ließ eine Serie von 10 Gewichten mit dem kleinsten Wert von 200 gr und jedem folgenden Wert $a_i = a_{i-1} + 1/10\ a_i$ über sieben Kategorien von "leicht" bis "schwer" beurteilen. In einem Fall ließ er das minimale Element der Serie (Experiment I), im anderen das maximale Element (Experiment II) mit einem "Wert" (value) assoziieren. In beiden Fällen wurde das Beurteilungsergebnis mit dem einer neutralen Serie verglichen. Tab. 9 zeigt die durchschnittliche Beurteilung für jedes Gewicht. Die Gewichte werden in Richtung des wertassoziierten Endes der Serie deutlich überschätzt, am entgegengesetzten Ende unterschätzt.

Element		1	2	3	4	5	6	7	8	9	10
Exp.I	W	2.00	2.49	2.87	3.46	4.01	4.55	5.16	5.61	6.29	6.64
	N	2.24	2.64	2.97	3.59	3.99	4.54	5.19	5.56	6.14	6.50
Exp.II	W	1.67	2.16	2.57	3.03	3.59	4.12	4.91	5.42	5.99	6.50
	N	1.92	2.38	2.69	3.07	3.64	4.36	4.81	5.41	5.96	6.43

Tab. 9. Experiment TAJFEL:
W = wertassoziierte Serie, N = neutrale Serie.

Das Ergebnis dieses Experiments entspricht denen sonstiger Experimente bei Gewichtung der Enden von über ein Merkmalskontinuum abgestuften Serien von Gegenständen. Die Gewichtung hat zur Folge, daß das gewichtete Ende der Serie in irgendeiner

Hinsicht überschätzt, das entgegengesetzte unterschätzt wird. Das Experiment von TAJFEL wurde hier ausgewählt, um zu zeigen, daß auch Wertungen "Anker"-Effekte bilden können, was für unsere Untersuchungen der interessantere Fall ist. Denn er bietet die Möglichkeit einer Erklärung für teilweise deutliches Übergewicht der "gut"-Urteile, die also nicht eine Folge einer schiefen Verteilung der beurteilten Gegenstände auf dem betrachteten Merkmalskontinuum zu sein braucht. Allerdings wären zusätzliche Interpretationsansätze dann notwendig, um zu erklären, warum dieses Übergewicht nicht in allen Fällen auftritt.

Unter der Voraussetzung, daß in den von uns untersuchten Fällen das Gegenstands-System relevante Bezugsgruppe ist für die Beurteiler, die Beurteiler selbst Mitglieder dieser Gruppe sind, die Relevanz der Bezugsgruppe auch oder vor allem sich auf das betrachtete Merkmalskontinuum bezieht, läßt sich von der Einstellungs- (attitude-) Forschung her ein weiterer Interpretationsansatz anbringen, wenn man bedenkt, daß das jeweils zur Diskussion stehende Merkmal (aus dem Leistungsbereich) für die verschiedenen Gruppenmitglieder möglicherweise eine verschiedene Relevanz besitzt. D.h. die Beurteilung eines Mitglieds hinsichtlich eines Merkmals hängt ab von der Position, die ein Gruppenmitglied jeweils auf dem Merkmalskontinuum einnimmt (es braucht hierbei die Richtung der Beziehung Relevanz - Position , letzteres heißt hier Leistungsausstattung, nicht diskutiert zu werden, den Fall der Wechselwirkung eingeschlossen).

SHERIFF et al. (25) fanden im Bereich der Einstellungsforschung, daß "... persons who are strongly committed to a stand on a controversial issue tend to use fewer categories and to distribute their judgments differently then persons who are less concerned with the issue " (S. 93). An anderer Stelle (SHERIFF & HOVLAND (24), S. 129) :

> " ... The degree of the individuals involvement in an issue should be closely related to important characteristics of his latitudes of acceptance and rejection. ... We summarized a study in which strongly ego-involved subjects were instructed to use eleven categories for judgment of favorableness - unfavorableness. Despite the instructions, they neglected certain categories, and the distribution of their judgments was closely related to their agreement or disagreement with the positions which the categories represented. These highly ego-involved subjects placed a large number of the items in the extreme categories with which they strongly disagreed and relatively fewer in the extreme categories with which they agreed. When given the opportunity to sort the same items into the number of categories they thought necessary, the highly involved subjects used fewer categories then less involved subjects and continued to pile up items at the end of the scale they rejected. From these results we infer that the highly ego-involved subjects had a more constricted latitude of acceptance and a more extensive latitude of rejection then the less involved subjects."

Tab. 10 zeigt diesen Sachverhalt deutlich: die "highly involved persons" haben zum einen ihre Einstellung zum vorgegebenen Sachverhalt praktisch reduziert auf Annahme und Ablehnung, zum anderen ist der Bereich der Ablehnung deutlich umfangreicher sowohl gegenüber dem Annahme-Bereich wie auch gegenüber dem Ablehnungs-Bereich der "less involved persons".

	latitudes of acceptance %	rejection %	noncommitment %
highly involved p.	36.9	59.2	3.9
less involved p.	40.3	38.6	21.1
p (diff.)	0.05	0.001	0.001

Table: Percentage of Statements on Reapportionment Accepted, Rejected, and Not Evaluated by Highly Involved and Less Involved Subjects.

Tab. 10. Aus SHERIFF et al. (25, S. 56).

In den von uns untersuchten Gruppen finden wir in einigen Fällen (s. Tab. 11) diesen von SHERIFF et al. gefundenen Sachverhalt wieder. Die jeweils sieben bzw. acht Besten einer Gruppe benutzen (bei der Zweier- und Dreier-Kategorisierung) die "schlecht"-Kategorie häufiger als die "gut"-Kategorie, der "mittel"-Bereich verschwindet nahezu. Umgekehrt verhalten sich die sieben oder acht Schlechtesten dieser Gruppen. Bei der Fünfer-Kategorisierung werden die beiden Extrem-Kategorien von den Besten deutlich stärker besetzt als von den Schlechte - sten(s. Abb. 13).

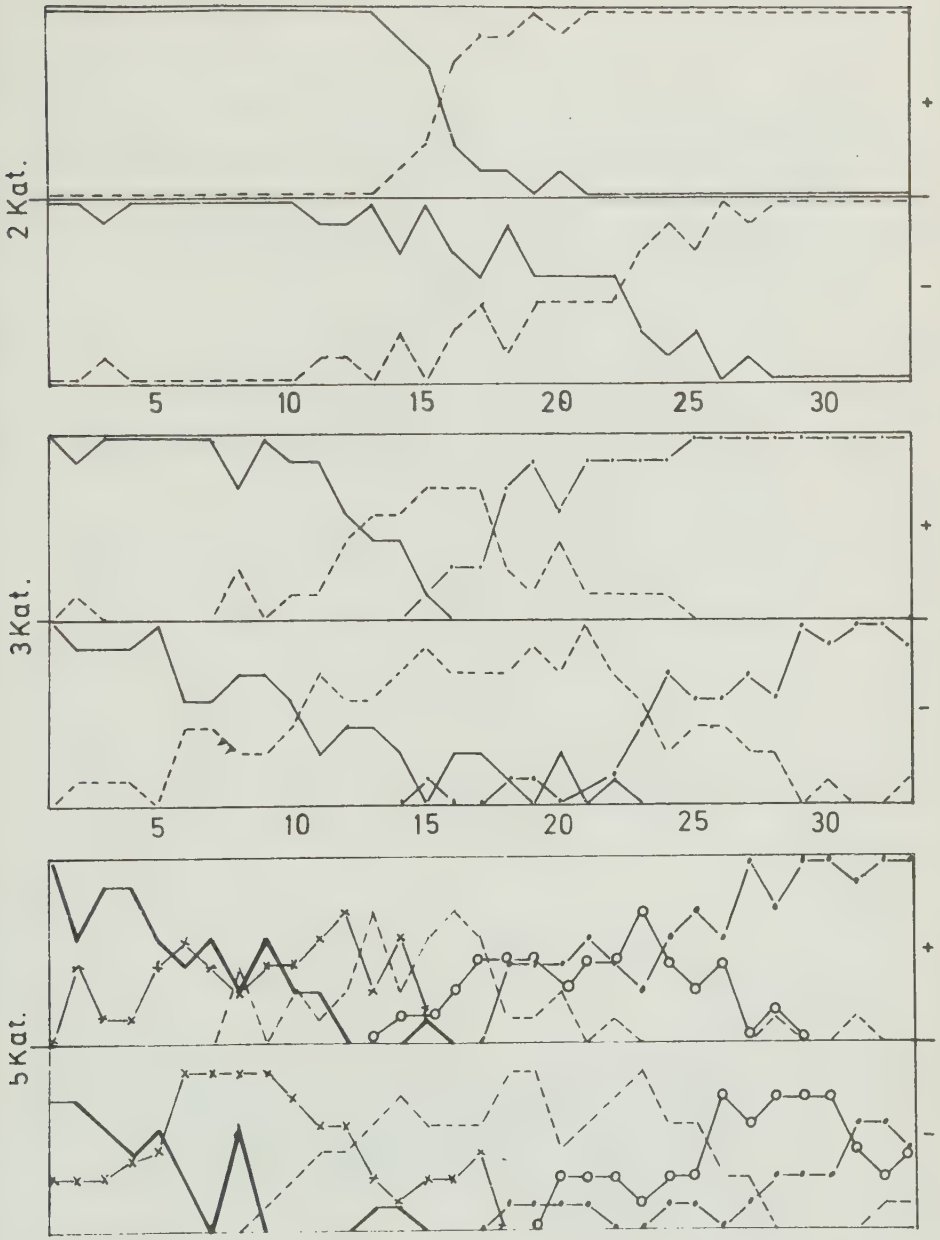


Abb. 13

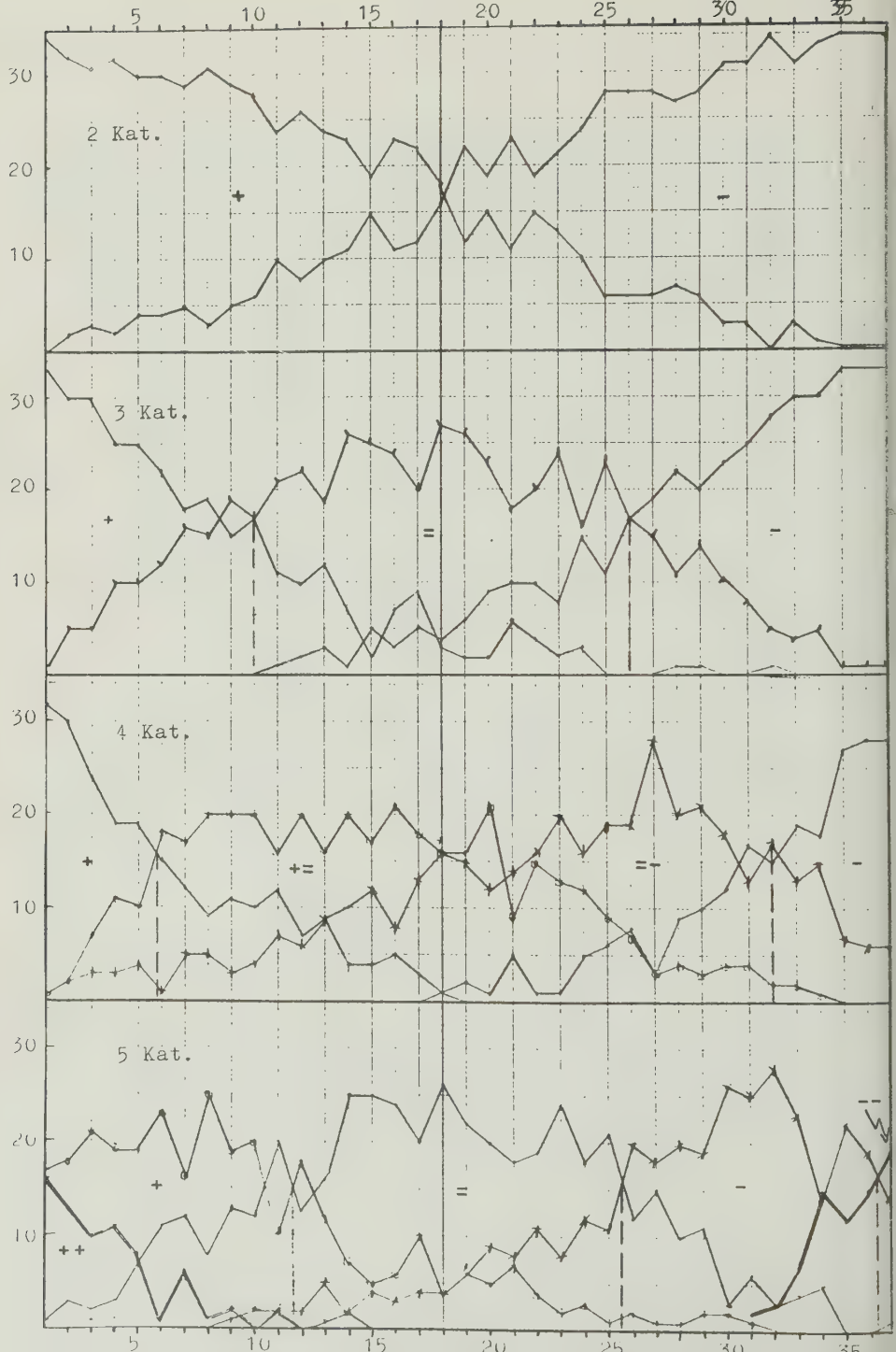
" + " gleich Beurteilungen durch die Klassenbesten.
" - " gleich Beurteilungen durch die Klassenschlechtesten.

Systemhöhe		2		3			5				
Beurteilung		+	-	+	=	-	++	+	=	-	--
Klasse	Beurteiler										
II	+	141	152	82	125	89	30	55	116	66	29
	-	168	128	108	139	49	27	85	108	53	23
III	+	134	202	78	93	165	22	68	71	78	96
	-	165	171	102	159	75	25	76	125	71	38
IV	+	129	144	92	71	110	49	58	46	65	55
	-	148	125	91	101	81	40	75	62	63	33
V	+	107	124	81	43	107	43	40	36	35	77
	-	136	95	72	97	62	27	54	82	42	26
VI	+	114	145	75	91	93	26	56	83	69	25
	-	144	115	68	122	69	11	78	99	55	16
VII	+	107	124	64	94	73	36	48	56	48	43
	-	111	120	69	86	76	52	40	64	36	39

Tab. 11. Anzahl der Urteile pro Kategorie der jeweils sieben bzw. acht am besten (+) bzw. am schlechtesten (-) beurteilten Beurteiler über die Systemhöhen (Differenzierungsgrad nach Anzahl der vorgegebenen Kategorien) 2, 3, 5. Die Beurteilungen bedeuten:
++ sehr gut, + gut, = mittel, - schlecht,
-- sehr schlecht.

Diese Befunde sind allerdings auch wieder nicht durchgängig (s. Tab. 11, Klasse VII und dazugehörig Abb. 14, die die Verteilung der kategorialen Urteile aller Mitglieder derselben Klasse wiedergibt). Die Entsprechung zu den SHERIFFschen ergibt sich offensichtlich nur dort, wo bei Berücksichtigung der Urteile aller Mitglieder einer jeweiligen Gruppe sich kein klares Bild der Verteilung der kategorialen Urteile (etwa Normalverteilung) ergibt. Da die kategorialen Urteile der am schlechtesten beurteilten Mitglieder einiger Gruppen sich annähernd normal verteilen, läßt sich möglicherweise die Behauptung aufstellen, daß in Gruppen mit Normalverteilung der Urteile das betrachtete Merkmal keine Relevanz besitzt bzw. in diesen Gruppen keine Untergruppen bestehen, für die das Merkmal erhöhte Relevanz besitzt.

Gibt es aber in Gruppen Untergruppen von systematisch divergierendem Charakter, ist es auch deshalb wiederum nicht zu erwarten, daß die Bedingungen metrischer System-Modelle erfüllt werden.


5
10
15
20
25
30
35
30
20
10
2 Kat.
+
-
3 Kat.
+
=
-
4 Kat.
+
+=
=-
-
5 Kat.
+
=
-
++
--

Klasse VII		Z-Werte				log-Werte			
		MIN	XQUER	MAX	MAX+1	MIN	XQUER	MAX	MAX+1
Lage d. Zäsuren nach d.empir.Skalenwerten									
Zäsur:									
1/2		1.171	-0.096	-0.096	-0.096	2.107	-0.169	-0.169	-0.169
1/3		0.554	-0.626	-0.753	-0.773	1.035	-1.092	-1.338	-1.397
2/3		1.781	0.531	0.531	0.531	3.169	0.923	0.923	0.923
1/4		0.345	-0.864	-0.953	-0.974	0.694	-1.511	-1.702	-1.751
2/4		1.171	-0.096	-0.096	-0.096	2.107	-0.169	-0.169	-0.169
3/4		2.168	0.917	0.932	0.936	3.860	1.615	1.646	1.655
1/5		0.000	-1.266	-1.291	-1.297	0.000	-2.273	-2.345	-2.345
2/5		0.878	-0.398	-0.398	-0.398	1.595	-0.694	-0.694	-0.694
3/5		1.634	0.384	0.384	0.384	2.917	0.664	0.664	0.664
4/5		2.472	1.009	1.299	1.299	4.406	1.789	2.286	2.434
Systemwerte:									
Zäsur:	erwartet:								
1/4	-25.0	-22.1	-24.1	-21.6	-21.5	-22.2	-24.2	-21.6	-21.4
3/4	25.0	27.2	27.4	27.2	25.7	27.7	27.2	27.6	27.7
1/5	-30.0	-31.6	-34.2	-30.3	-30.0	-32.7	-37.8	-30.7	-30.0
2/5	-10.0	- 7.9	- 8.9	- 7.7	- 7.3	- 7.9	- 9.4	- 7.5	- 7.1
3/5	10.0	12.8	12.5	12.5	12.5	12.6	12.6	12.6	12.6
4/5	30.0	35.5	29.4	36.1	36.9	36.0	30.0	37.4	36.9

Tab. 12. Zäsurenwerte.

(g) Mitgliedsgruppe - Bezugsgruppe.

Als Mitgliedsgruppe kann diejenige Gruppe aufgefaßt werden, zu der jemand gehört aufgrund seiner Berufszugehörigkeit, aufgrund organisatorischer Maßnahmen etwa in Betrieben, Sportvereinen, in der Armee usw. oder aufgrund sonstiger Bedingungen. Sie ist die Gruppe, in der er mit den anderen Mitgliedern in irgendeiner Art in Interaktion steht und/oder zu der er sich als zugehörig fühlt und/oder zu der er zugeordnet ist und sich zugeordnet weiß. Sie ist also die Gruppe einer Zugehörigkeit aufgrund irgendwelcher Bedingungen tatsächlicher Art und beschreibt einen Ist-Zustand. Dagegen kann die Bezugsgruppe, und wird auch manchmal in der Literatur (z.B. 9), als die Gruppe aufgefaßt werden, deren Einstellung, Verhaltensweisen, Normen, Leistungsfähigkeit für die eigenen Einstellungen, Verhaltensweisen, Normen und Einschätzungen der eigenen Leistungsfähigkeit relevant sind. Mitgliedsgruppe und Bezugsgruppe können identisch sein, sie können aber auch hinsichtlich ihrer Mitglieder unterschiedlich sein. Das bedeutet im letzteren Fall die tatsächliche Angehörigkeit zu einer Mitgliedsgruppe bei Relevanz, in irgendeiner Hinsicht, einer anderen, "elementenfremden" Gruppe, der Bezugsgruppe. Ist- und Soll-Zustand fallen auseinander.

Die in der Literatur anzutreffende Annahme des Auseinanderfallens der hinsichtlich ihrer Mitglieder elementfremden Mitglieds- und Bezugsgruppe läßt sich allerdings (vermutlich in allen Fällen) bei genauerem Hinsehen schwerlich auf-

rechterhalten. Dagegen kann von Fall zu Fall verschieden nicht ausgeschlossen werden, bzw. kann nachgewiesen werden, daß, neben dem trivialen Fall der Identität beider Gruppen, die eine Gruppe Untergruppe der anderen ist, bzw. hinsichtlich der Menge der Mitglieder die eine als Teilmenge in der anderen enthalten ist. Diese Relation des Enthaltenseins ist für beide Gruppen zueinander möglich. Also in formaler Auffassung:

a) M (B
b) B (M ,

wobei M = Mitgliedsgruppe
B = Bezugsgruppe
(= Relation des Enthaltenseins.

Der Fall (a) kann dann eintreten, wenn jemand Mitglied in zwei oder mehr Gruppen ist, die experimentelle Befragung aber nur eine Gruppe erfaßt, und die Gruppen hinsichtlich des befragten Merkmalskontinuums eine Gemeinsamkeit aufweisen und mindestens eine Gruppe hinsichtlich dieses Merkmals ein mehr oder weniger. Bei uns hieße das z.B., daß jemand Mitglied einer Schulklasse ist und gleichzeitig eines Sportvereins, so daß das zum Bezugsystem gehörige Gegenstands-System diese beiden Gruppe umfaßt, die experimentelle Befragung aber nur die eine Gruppe (hier die Schulklasse) zur Beurteilung vorgibt. Die Bezugsgruppe braucht aber nicht zu bestehen aus allen den Gruppen, zu denen jemand tatsächlich gehört. Virtuelle Gruppen (oder auch nicht Gruppen, sondern einzelne Individuen), angestrebte Gruppen oder aufgrund irgendwelcher Zusammenfassungen ("alle gleichen Alters", "alle überhaupt Bekannten" oder sonstwie) gebildete Gruppen können neben der betrachteten Mitgliedsgruppe als zu einer

Bezugsgruppe zugehörig aufgefaßt werden.

Der Fall (b) ergibt sich dadurch, daß die Zugehörigkeit zu einer Mitgliedsgruppe mehrfach bedingt sein kann, unter Umständen nur "äußerliche" organisatorische Gründe hat (wie meist bei Schulklassen), so daß dann über ein herausgegriffenes Merkmalskontinuum nicht alle Mitglieder als zur Bezugsgruppe zugehörig erlebt werden.

Die Definition einer Bezugsgruppe als einer Gruppe, die elementfremd zu der gegebenen Mitgliedsgruppe ist und deren Zugehörigkeit in irgendeiner Weise angestrebt wird, betrifft entweder nicht das zu einem Bezugssystem zugehörige Gegenstands-System, oder, wird sie als ein solches aufgefaßt, muß die gegebene Mitgliedsgruppe als zu ihr zugehörig aufgefaßt werden. Es wäre dann jeweils aufzuweisen, welches das zugehörige Merkmalskontinuum ist, wenn man den Begriff des Bezugssystems aufrechterhalten will und nur dann von einem Bezugssystem die Rede sein kann, wenn es sich über ein solches aufspannt.

Die Divergenz von Mitglieds- und Bezugsgruppe hinsichtlich der zugehörigen Mitglieder kann dazu führen, und führt auch im allgemeinen dazu, daß die Bedingungen metrischer System-Modelle nicht erfüllt werden. Eine jeweils betrachtete Gruppe ist zunächst nur Mitgliedsgruppe. Inwieweit diese Gruppe auch Bezugsgruppe ist, läßt sich nicht von vornherein behaupten. Die Feststellung einer Bezugsgruppe ist aber notwendig für die Entdeckung einer Isomorphie der metrischen Struktur zu einem System-Modell. Die Feststellung einer Bezugsgruppe dürfte

Systemhöhe	Kategorie	Gruppe I Summe=314	Gruppe II Summe=276
2	+	259	194
	-	55	82
3	+	154	142
	=	123	107
	-	27	27
4	++	138	104
	+	116	103
	-	40	53
	--	20	16
5	++	68	45
	+	116	93
	=	97	101
	-	21	31
	--	12	6

Tab. 13. Anzahl der Urteile pro Systemhöhe und Kategorie. Es bedeuten: ++ sehr gut, + gut, = mittel, - schlecht, -- sehr schlecht.

im Einzelfall allerdings experimentell außerordentlich schwierig sein, wenn nicht gar unmöglich. In den in Tab. 13 gezeigten Beispielen läßt sich allerdings mit einigen guten Gründen zumindest vermuten, zwar nicht, welche die Bezugsgruppe ist und wie das gesamte Bezugssystem aussieht, aber doch, daß die Mitgliedsgruppe nicht Bezugsgruppe ist. Es handelt sich hierbei um eine Gruppe von Jugendlichen eines Sport-Vereins, die in einer Liga gegen altersgleiche Jugendliche anderer Vereine regelmäßig Mannschaftsspiele (hier Fußballspiele) durchführt und mit derselben Regelmäßigkeit (im Sprachgebrauch der Sportberichterstattung) "Kantersiege" erreichte. Stellen die zur selben Liga gehörigen Gruppen aller Vereine insgesamt die Bezugsgruppe dar, so ist es einsichtig, daß die hier befragten Jugendlichen sich selbst gegenseitig am oberen Ende des Leistungskontinuums einordnen.

8. Beurteilungen über ein nicht anschauliches Leistungsmerkmal.

In Abb. 15 stellen wir die Ergebnisse von Beurteilungen aller über alle innerhalb einer Schulklasse von im Durchschnitt fünfzehnjährigen männlichen Jugendlichen dar. Abweichend von den vorangehend gezeigten Ergebnissen handelt es sich hier um ein Merkmalskontinuum aus dem Leistungs- oder Befähigungsbereich, das nicht wie im sportlichen Bereich unmittelbar anschaulichen Charakter besitzt. Die Jugendlichen waren aufgefordert worden, sich hinsichtlich ihrer "Lernfähigkeit" zu beurteilen und zwar nach folgenden Kategorien (dieselben wie sonst):

2 Kategorien: gut, schlecht;
3 " : gut, mittel, schlecht;
4 " : gut, eher gut, eher schlecht, schlecht;
5 " : sehr gut, gut, mittel, schlecht, sehr schlecht.

Der Begriff "Lernfähigkeit" wurde umschrieben mit: wie gut einer lernen kann, und zwar nicht nur auswendig, wie gut einer etwas begreifen, verstehen kann u.ä., also einer allgemeinen Leistungsfähigkeit in schulischen Anforderungen. Der Begriff wurde ersatzweise eingeführt für in der Schule gezeigte allgemeine intellektuelle Leistungsausstattung, Leistungsfähigkeit. Der Begriff der Intelligenz wurde vermieden als möglicherweise zu affektbesetzt und deshalb die Ergebnisse verzerrend. Nicht zuletzt wurde er auch deshalb nicht verwendet, um ebenfalls das Ergebnis möglicherweise verzerrende gruppendynamische Prozesse zu

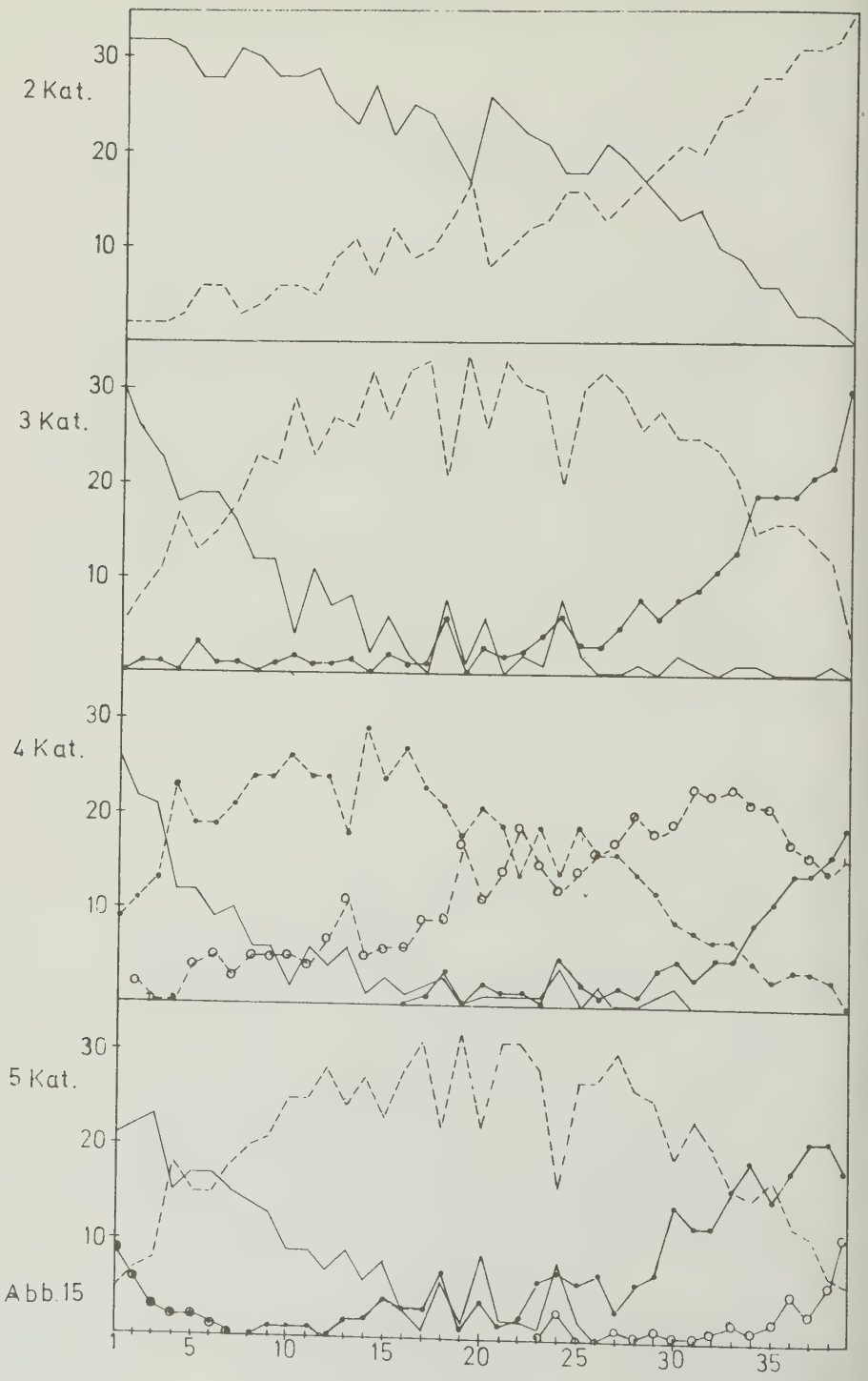


Abb.15

vermeiden. Da der Untersucher einmal der Klasse unbekannt war, zum anderen durch einen Lehrer eingeführt werden mußte, konnte es von seiten der Schüler nicht ausgeschlossen werden, daß er mit dem Lehrer "unter einer Decke steckte". Da außerdem die Beurteilung der Intelligenz für den gesamten Schulerfolg eines Schülers erhebliche Bedeutung hat (im Unterschied zu der im Rahmen einer Schule relativ irrelevanten Leistungsfähigkeit in einer Sportart), erschien es nicht ausgeschlossen, daß es zu einer Solidarisierung der Klassenmitglieder gekommen wäre und zwar in der Form, daß jeder die anderen gleichmäßig eingestuft hätte: entweder alle als "gut" oder alle als "mittel".

Das Ergebnis verifiziert offensichtlich die Normalverteilungs-Hypothese für Beurteilungen jedenfalls im Bereich der Leistungskontinua. Es läßt sich zwar nicht ausschließen, daß der mächtige "mittel"-Bereich teilweise bedingt ist durch die oben dargestellten Solidarisierungs-Prozesse. Die Tatsache jedoch, daß die "gut"- und "schlecht"-Bereiche ebenfalls deutlich ausgeprägt sind, spricht allerdings nicht für diese Vermutung. Ebenfalls nicht für eine mögliche Vermutung, daß der Begriff "Lernfähigkeit" zu diffus ist, so daß zur Vermeidung grober Mißbeurteilungen einfach die mittlere Kategorie bevorzugt verwendet wurde.

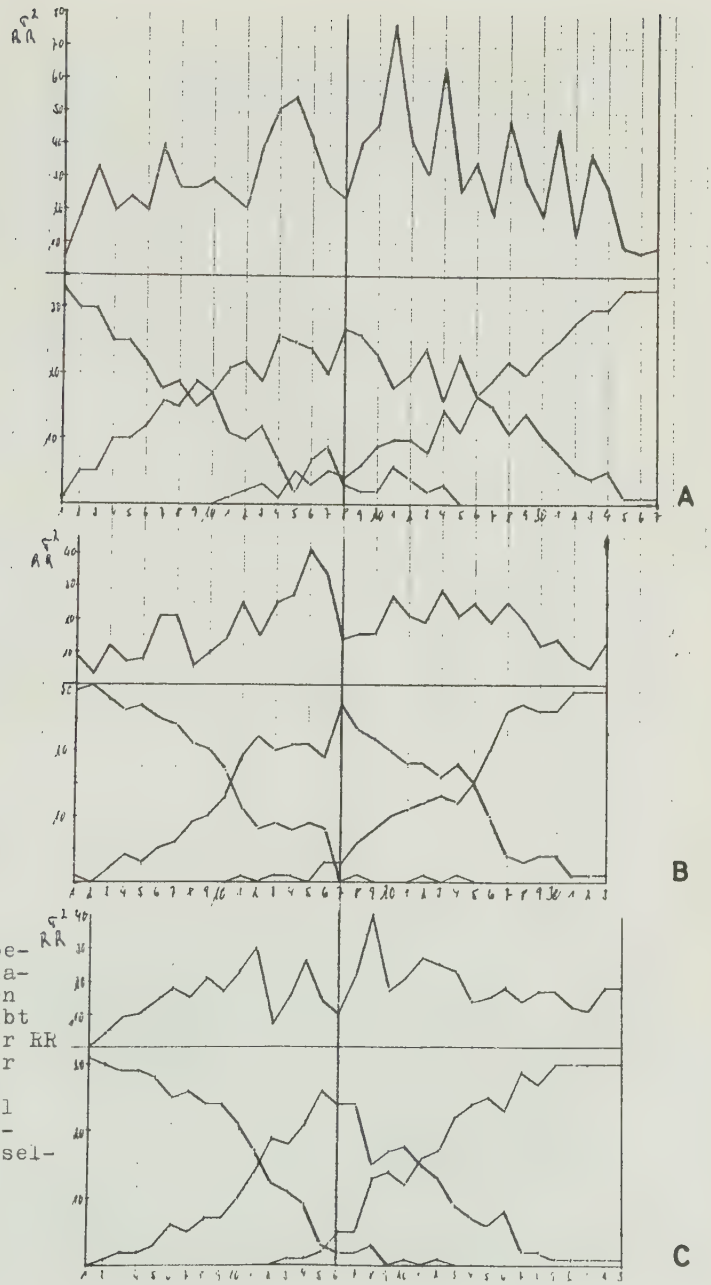
Ein Vergleich der Lagen der Zäsuren zeigt, daß es sich hier vermutlich um ein dreistufiges System handelt. Denn die den "mittel"-Bereich berandenden Zäsuren haben im dreistufigen wie im fünfstufigen

System dieselbe Lage (was allein freilich noch kein ausreichendes Argument ist), zum anderen stellen im fünfstufigen System die Extrembereiche unselbständige Bereiche dar (was vom dreistufigen System her zu erwarten war).- Im übrigen haben wir bei dieser Klasse - abweichend von der Vorgehensweise bei den anderen Klassen - die Kategorisierungen nach 2,3,4 und 5 Kategorien in bunter Reihenfolge - d.h. von beurteilender Person zu Person in anderer Abfolge - vornehmen lassen.

9. Vergleich Kategorien-Skalen - Rangreihen.

Neben der gegenseitigen Beurteilung aller Mitglieder der Klassen aufgrund vorgegebener Eigenschaften-Systeme haben wir Rangreihen (RR) hinsichtlich desselben Merkmalskontinuums erstellen lassen. D.h. jedes Mitglied einer Klasse erstellte über dasselbe Merkmalskontinuum wie bei den Kategorisierungsversuchen (aber in zeitlicher Reihenfolge erst nach diesen) über alle anderen der Klasse (also sich selbst ausgeschlossen) eine Rangreihe beginnend mit dem Besten, der den Rangplatz 1 erhalten sollte.

Sollte es sich bei einem gegebenen Bezugsystem um ein stabilisiertes System handeln, so war die Vermutung, daß diese Stabilisierung auch in Rangreihen nachzuweisen ist. Dieser Nachweis ließe sich erbringen über die Ähnlichkeit struktureller, für Bezugsysteme typische Momente. Gliedert sich ein Bezugsystem (nach dem WITTE-schen Modell) von dem in irgendeiner Hinsicht



Der jeweils obere Teil der Diagramme in allen drei Fällen gibt die Varianz der RR pro beurteilter Person wieder, der untere Teil die 3-Kategorisierung in derselben Klasse.

Abb. 16

minimalen und maximalen Element der Trägermenge T des Gegenstands-Systems $\mathcal{T}$ (der Pole) her über ab-(zu-)nehmende Ähnlichkeit auf dem betrachteten Merkmalskontinuum von Pol zu Pol, so heißt das von der empirischen Zuordnung her, daß die Pol-Elemente eindeutig bzw. überwiegend einer Eigenschaft (Kategorie) des Eigenschaften-Systems $\mathcal{E}$ und zwar der jeweils irgendwie minimalen oder maximalen Eigenschaft zugeordnet werden. Neben diesen beiden hervorragenden Örtern stellt aus strukturgenetischen Gründen (hierzu siehe WITTE (33, 34)) die System-Mitte ebenfalls einen hervorragenden Ort dar. D.h. auch hier müßte die Zuordnung eindeutig bzw. überwiegend sein im gegebenen Fall ("im gegebenen Fall" soll heißen, daß natürlich experimentell die Möglichkeit einer Zuordnung gegeben ist, also in allen Fällen ungeradzahliger Mächtigkeit der Menge der Eigenschaften des vorgegebenen Eigenschaften-Systems (3,5,7 Kategorien)).

Als Maß des Grades dieser Eindeutigkeit (streng genommen natürlich ein Unbegriff) der Zuordnung der Eigenschaften zu den Elementen der Trägermenge kann man die Streuung der in eine Rangreihe gebrachten und natürlichen Zahlen zugeordneten Eigenschaften über ein jeweiliges Element annehmen. Bei Eindeutigkeit der Zuordnung wäre also die Streuung s = 0. Die oben angegebene Vermutung bedeutet also, daß mit der Eindeutigkeit der kategorialen Zuordnung eine Eindeutigkeit des Rangplatzes verbunden ist.

Tatsächlich läßt sich in den in Abb. 16 gezeigten Fällen dieser Sachverhalt aufweisen. Die an-

gegebenen ausgezeichneten Örter des Systems sind auch in den Rangreihen ausgezeichnete Örter, d.h. die Rangplatzzuweisung weist hier eine geringe Streubreite auf. Die Korrelationen der Varianzen der kategorialen Zuordnung pro Element der Trägermenge mit der der RR-Zuordnung ergeben für $r_A = 0.73$, $r_B = 0.71$, $r_C = 0.68$. Daß auch die Mitte ein derartig ausgezeichneter Ort ist, zeigt sich daran, daß pro untersuchter Klasse die Korrelationen für

	Anzahl der Kategorien			
Klasse	3	4	5	6
II	0.51	0.43	0.58	0.52
III	0.32	0.14	0.06	0.23
IV	0.49	0.59	0.64	0.44
V	0.68	0.58	0.57	0.56
VI	0.73	0.63	0.39	/
VII	0.71	0.66	0.85	/

Tab. 14. Korrelationen der Varianzen der pro Element der Trägermenge zugeordneten Kategorien mit den Varianzen der pro Element zugeordneten Rangplätze.

jeweils ungeradzahlige Anzahlen der Eigenschaften pro Eigenschaften-System am höchsten sind (s. Tab.14). Sie wären möglicherweise höher, wenn man den Sach-

verhalt berücksichtigt hätte, daß in einzelnen Fällen die Mitten bei den verschiedenen Untergruppen verschieden gelegen sind.

Weitergehende Analysen des Zusammenhangs zwischen Kategorien-Skalen und Rangreihen sind hier allerdings sowieso nicht vorgenommen worden. Die Erhebung der Rangreihen geschah nur aus Gründen des Experiments im Sinne des "Herumprobierens". Die kurze Darstellung kann aber vielleicht Anlaß geben für Spezialuntersuchungen über diesen Zusammenhang.

A N H A N G

Anhang I. (S.Abb. 2)

Auf 28 Karten weißen Kartons vom Format DIN A 7 waren 27 rote und/oder grüne Punkte von ca. 0.5 cm Durchmesser willkürlich verteilt wie folgt angebracht:

	Anzahl der Punkte								
Nr.	1	2	3	4	.....	25	26	27	28
rot	27	26	25	24	.....	3	2	1	0
grün	0	1	2	3	.....	24	25	26	27
Summe	27	27	27	27	.....	27	27	27	27

Für die Beurteilung hinsichtlich der Farbe wurden folgende Kategorien verwendet:

2 Kategorien: rot, grün;
3 " : rot, mittel, grün;
4 " : rot, eher rot, eher grün, grün;
5 " : rot, mittel rot, mittel, mittel grün, grün.

Die Vorlagen wurden kurzzeitig dargeboten.
10 Versuchspersonen.

Anhang II. (S. Abb. 3)

Beurteilt wurden 30 Striche von 0.8 mm Breite und der Länge von 1, 2, 3, ... , 30 cm. Die Striche standen in 1 cm Abstand in bunter Reihenfolge auf einer Grundlinie alle in einer Richtung zur Versuchsperson hin (die ursprüngliche Absicht, die mit dieser Anordnung verbunden war, braucht hier nicht zu interessieren). Für die Beurteilung der Striche hinsichtlich ihrer Länge wurden folgende Kategorien verwendet:

2 Kategorien: groß, klein;
3 Kategorien: groß, mittel, klein;
4 Kategorien: groß, eher groß, eher klein, klein;
5 Kategorien: groß, mittel groß, mittel, mittel klein, klein.

12 Versuchspersonen.

Anhang III (PFANZAGL-Versuch).

Der Versuch wurde in der Klasse durchgeführt, deren Ergebnisse der Güte-Kategorisierung in Abb. 8 dargestellt sind. Um zu einer zwischen den Beurteilern unabhängigen Messung der Systemmitte zu gelangen, wurde die Klasse in zwei Gruppen geteilt. Die Gruppe I führte folgende Operationen der Mittenbildung durch:

1.	P_o		M_1	P_1
2.	P_o	M_2'	M_1	
3.			M_1	M_2'' P_1

Die Gruppe II:

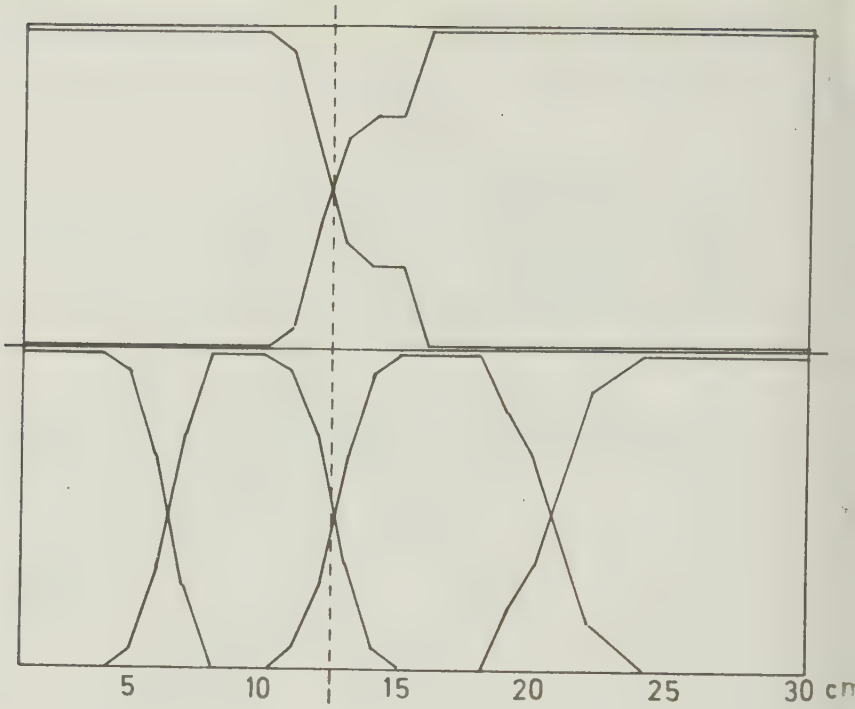
4. M_2' M_1' M_2''

Die Forderung des Bisymmetrie-Axioms von PFANZAGL besteht dann darin, daß $M_1 = M_1'$. Die eben angegebene experimentelle Anordnung entspricht allerdings nicht vollständig dem Axiom.

Die Ergebnisse waren wie folgend:

Beurteilte Pole und Mitten	Rangreihe der beurteilten Mitglieder (wie in Abb. 8) 1234567890123456789012345678901234567890123456789
	x x x
P_o P_1	925131 3 911 52 6
M_1	1 1 131121233 11 111
M_2'	1 1175 22 2 1 1 1
M_2''	3 11 31 1 14115 1 1
M_1'	323 1 1 11

Die Tabelle zeigt die Urteilshäufigkeiten für die verschiedenen Mitten und Pole à beurteilter Person. Die Zeichen (x) unter der Rangreihe zeigen die Lage der Zäsuren 1/4, 1/2, 3/4. Es ergibt sich weder eine Übereinstimmung der Mitten M_1 und M_2', noch eine Übereinstimmung beider mit der Lage der Zäsur 1/2 (=2/4). Von der Häufigkeitsverteilung der Urteile her kann man allerdings hier kaum von einer Mitten-Bildung sprechen. Für M_1 wurden 24 Urteile auf 16, für M_1' 12 Urteile auf 7 Klassenmitglieder verteilt. Der Versuch muß deshalb insgesamt als gescheitert angesehen werden.


5
10
15
20
25
30 cm

| | | |

Wir haben einen weiteren PFANZAGL-Versuch an einer Strich-Menge neben einem Kategorisierungsversuch vorgenommen. Die voranstehende Abbildung zeigt die Ergebnisse des Kategorisierungsversuchs mit zwei (klein, groß) und vier Kategorien (klein, eher klein, eher groß, groß) sowie die Verteilungen der Urteile zu den einzelnen Mitten.

Arithmetische Mittel: M_1 = 16.82 cm
M_1' = 14.82 cm
M_2' = 9.45 cm
M_2'' = 21.73 cm

Zäsur (1/2, 2/4): Z = 12.5 cm

Strichlängen: AM = 15.5 cm
GM = 12.45 cm

Auch hier streuen die Urteile bei den verschiedenen Mittenbildungen stark. Die beiden Mitten M_1, M_1' kommen nicht zur Deckung; beide liegen überhalb der Zäsur (1/2, 2/4). Die ersteren liegen um das arithmetische Mittel (AM), die Zäsur um das geometrische Mittel aller Strichlängen !

Varianzanalyse (M_1, M_1', Z. 2/4):

Var.quelle	SAQ	FG	Var.	F	p
innerhalb	319	34	9.4		
zwischen	149	2	74.5	7.92	0.01
gesamt	468	36			

Die Unterschiede der durchschnittlichen Lage zwischen den drei angegebenen Mitten sind also als bedeutsam anzusehen.

Anhang IV (s. Abb. 10).

Das Experiment war gedacht als Modell-Experiment unter kontrollierten Bedingungen auf seiten der Trägermenge zur Frage der Normalverteilung kategorialer Bereiche.
Auf einem weißen DIN A 4 Blatt wurden 39 Striche bunt verteilt. Der kürzeste Strich hatte eine Länge von 1 cm, der längste eine von 20 cm. Die Länge der übrigen Striche wurde so konstruiert, daß die Fläche unter der Normalverteilung in 39 gleiche Teile aufgeteilt wurde. Der Ort auf der x-Achse unter der Normalverteilung, der der Fußpunkt der Flächenhalbierenden des jeweiligen Flächensektors war, ergab den Wert in u(s) für den jeweiligen Strich. Der in cm mittelste Strich lag dadurch bei 0.00 u(s). Die so erhaltenen Werte in u(s) wurden linear auf das Intervall [1,20] transformiert, wodurch sich die Länge der Striche der betrachteten Strichmenge in cm ergab. Die Abstände in cm benachbarter Striche der geordneten Strichmenge werden also zur Mitte hin zunehmend geringer. Die nachfolgende Tabelle gibt die Länge der Striche der betrachteten Strichmenge in cm und die zugehörigen u-Werte wieder:

Nr.	cm	Nr.	cm	± u(s)
1	1.0	39	20.0	2.230
2	4.4	38	16.65	1.77
3	5.4	37	15.6	1.52
4	6.0	36	15.0	1.345
5	6.5	35	14.5	1.20
6	6.95	34	14.1	1.075
7	7.33	33	13.67	0.97

(Forts. Tab. S. 122 :)

Nr.	cm	Nr.	cm	± u(s)
8	7.65	32	13.35	0.87
9	7.95	31	13.1	0.78
10	8.2	30	12.8	0.695
11	8.5	29	12.5	0.615
12	8.75	28	12.25	0.54
13	9.o	27	12.0	0.47
14	9.2	26	11.8	0.4
15	9.45	25	11.5	0.33
16	9.7	24	11.3	0.26
17	9.9	23	11.1	0.195
18	10.1	22	10.9	0.13
19	10.3	21	10.7	0.065
20	10.5			0.000

Die nachfolgende Tabelle gibt die beobachtete und die (nach dem WITTEschen Modell) erwartete Lage der Zäsuren (in cm) wieder sowie die Anzahl der beurteilten Objekte pro kategorialem Bereich:

Zäsur	beobachtet cm	erwartet cm	n Objekte
1/2	10.5	10.5	20/19
1/3	7.4	7.33	7/25/7
2/3	13.5	13.66	
1/4	6.0	5.75	
2/4	10.5	10.5	3.5/16/16.5/3
3/4	15.3	16.25	
1/5	4.5	4.8	
2/5	8.5	8.6	1.5/8.5/17/
3/5	12.2	12.4	10/2
4/5	16.1	16.2	

Systemhöhe	2		3			4				5				
Kategorie	1	2	1	2	3	1	2	3	4	1	2	3	4	5
Klasse														
I	835	571	535	630	241	408	510	386	102	230	366	515	236	59
II	686	610	406	601	288	271	436	412	177	122	302	504	268	100
III	753	927	482	573	626	372	407	440	462	168	401	426	356	330
IV	686	758	454	455	535	352	372	332	388	190	299	358	345	252
V	532	492	329	309	384	261	263	216	284	131	220	253	195	225
VI	602	656	312	537	409	237	396	406	219	74	306	452	343	85
VII	481	511	304	379	309	232	290	262	208	216	194	229	151	202
VIII	798	567	253	874	238	168	611	451	135	25	254	802	246	38

Anhang V. Anzahl der Urteile pro Kategorie, Systemhöhe und Klasse.
Die Abfolge der Kategorien wie sonstig im Text.

Systemhöhe	6						7						
Kategorie	1	2	3	4	5	6	1	2	3	4	5	6	7
Klasse													
I	198	306	434	309	105	54	170	278	241	381	173	83	43
II	115	235	363	334	169	80	86	194	238	292	196	126	79
III	158	319	349	341	241	273	134	320	224	253	230	225	254
IV	188	238	295	259	250	214	175	242	185	229	283	221	109
V	125	188	209	145	154	203	120	181	122	139	93	131	206
VI,VII,VIII	entfällt												

Anhang V. Fortsetzung.

Literaturverzeichnis.

(1) Ahlert, Frohmut — Experimenteller Beitrag zur phänomenalen Gleichabständigkeit der Grauleiter, untersucht mit Transponierungs- und Mittelungsverfahren. Zulassungsarbeit, Tübingen, 1962, unveröffentlicht.

(2) Asch, S.E. — Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego-standards. J.Social Ps.,1940,12,433-465

(3) Atkins,A.L., Deaux,K.K., Bieri,J. — Latitude of acceptance and attitude change:Empirical evidence for a reformulation. J.Pers.Soc.Ps.,1967,6,47-54

(4) Deaux,K., Bieri,J. — Latitude of acceptance in judgments of masculinity - femininity. J.Pers.,1967,35,109-117

(5) Ekman,G., Sjöberg,L. — Scaling. Ann.Rev.Ps.,1965,16,451-474

(6) Eriksen,C.W., Hake,H.W. — Absolute judgments as a function of stimulus range and number of stimulus and response categories. J.Exp.Ps.,1955,49,323-332

(7) Fehrer,E. — Shifts in scale values of attitude statements as a function of the composition of the scale. J.Exp.Ps.,1952,44,179-188

(8) Festinger,L. — Wish,expectation,and group standards as factors influencing the level of aspiration. J.Abn.Soc.Ps.,1942,37,184-200

(9) Hartley,R.E. Relationship between perceived values and acceptance of a new reference group. J.Soc.Ps.,1960,51,181-190

(10) Heckhausen,H. Leistungsmotivation. Hdb.d.Ps.,Bd.II,602-702, Verlag f.Ps.,Göttingen,1965

(11) Heckhausen,H. Tübinger Symposion über Motivation: Eine Rahmentheorie der Motivation in zehn Thesen. Z.f.exp.ang.Ps.,1966,13,464-483

(12) Helson,H. Adaptation-Level Theory. Harper&Row,New York,1964

(13) Hilgard,E.R. et al. Level of aspiration as affected by relative standing in an experimental social group. J.Exp.Ps.,1940,27,411-421

(14) Hruschka,Erna Experimenteller Beitrag zur Struktur ein-dimensionaler Bezugsysteme mit Dingcharakter. Tübinger Dissertation,1959, unveröffentlicht.

(15) Johnson,D.M. et al. Systematic study of end anchoring and central tendency of judgment. J.Exp.Ps.,1964,67,501-506

(16) Katz,F.M. The meaning of success:some differences in value systems of social classes. J.Soc.Ps.,1964,62,141-148

(17) Kellner,B. Untersuchungen an bezugsystembildenden Mannigfaltigkeiten, die nicht auf physikalische Maßgrundlagen bezogen werden können. Zulassungsarbeit Münster,1966, unveröffentlicht.

(18) Miller,G.A. The magical number seven,plus or minus two: some limits on our capacity for processing information. Ps.Rev.,1956,63,81-97

(19) Parducci,A. et al. Context effects in judgment: Adaptation level as a function of the mean,midpoint,and median of stimuli. J.Exp.Ps.,1960,60,65-77

(207 Parducci,A. Range-frequency compromise in judgment. Ps.Monographs,1963,77

(21) Parducci,A. et al. Stimulusavaraging and the context for judgment. Perception & Psychophysics, 1968,3,145-150

(22) Pfanzagl,J. Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens. Würzburg:Physica Verlag,1959

(23) Sherif,M. et al. Assimilation and contrast effects of anchoring stimuli on judgments. J.Exp.Ps.,1958,55,150-155

(24) Sherif,M., Hovland,C.I. Social Judgment. New Haven:Yale University Press,1961

(25) Sherif,Carolyne, Sherif,M., Nebergall,R.E. Attitude and Attitude Change. Philadelphia and London,1965

(26) Sixtl,F. Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme. Weinheim(Bergstraße):Beltz,1967

(27) Sjöberg,L. A study of four methods for scaling paired comparisondata. Scand.J.Ps.,1965,6,173-185

(28) Suppes,P., Zinnes,J.L. Basic measurement theory. In: Handbook of Mathematical Psychology I,1-76. New York:Wiley,1967

(29) Tajfel,H. Value and the perceptual judgment of magnitude. Ps.Rev.,1957,64,192-204

(30) Thurstone,L.L. The Measurement of Values. Chicago:The University of Chicago Press,1960

(31) Torgerson,W.S. Theory and Methods of Scaling. New York:Wiley,1963

(32) Weber,Renate E. Experimentelle Untersuchung über die anschauliche Größe zweidimensionaler Gegenstände. Ein Beitrag zum Problem psychischer Bezugsysteme. Dissertation,Tübingen,1963

(33) Witte,W. Struktur,Dynamik und Genese von Bezugsystemen. Ps.Beiträge,1960,4,218-252

(34) Witte,W. Über Phänomenskalen. Ps.Beiträge,1960,4,645-672

(35) Witte,W. A mathematical model of reference systems and some implications for category scales. Acta Psychologica,1961,19,378-382

(36) Witte,W. Das Problem der Bezugsysteme. Hdb.d.Ps.,I,1.Halbbd.,1003-1027, Göttingen:Verlag f.Ps.,1966

(37) Orlik,P. Theoretische und experimentelle Beiträge zur Psychologie der Ähnlichkeit im Lichte der Bezugsystemforschung. Dissertation,Tübingen,1963

Danksagung

Herrn Prof. Dr. Wilhelm Witte danke ich für die Anregungen, die zu dieser Arbeit führten, und für zahlreiche kritische Hinweise.

Für Unterstützung der aufwendigen Arbeiten, die zur Erstellung der verwendeten Rechen-Programme notwendig waren, danke ich Herrn Dipl.-Psych. Langheinrich.

Den Rektoren und Lehrern der Wichernschule und Erich-Klausener-Schule in Münster bin ich zu Dank verpflichtet für ihr bereitwilliges Entgegenkommen, das die Durchführung der Versuche in ihren Schulen ermöglichte.

Lebenslauf

Ich wurde am 22.1.1939 in Hongkong/China als Sohn des Carl Walter Sievers (damals als Reedereikaufmann in China tätig) und seiner Ehefrau Irma, geb. Prahl, geboren.

Während der Internierung der Familie als Zivilgefangene der britischen Kolonialmacht in Indien besuchte ich von 1944-1947 bis zur 3. Klasse, sodann von 1947-1949 in Hamburg die Grundschule. Danach wechselte ich auf das altsprachliche Gymnasium "Christianeum" in Hamburg über und bestand dort 1958 das Abitur.

Von 1958-1959 hielt ich mich zunächst privat für knapp ein Jahr in den USA auf, um dann mit Beginn des Sommer-Semesters 1959 an der Universität Hamburg das Studium der Philosophie, der Alten Sprachen und der Kunstgeschichte zu beginnen, welches ich von 1960-1962 unter Einschluß der Soziologie an der Universität Tübingen fortsetzte. Im Winter-Semester 1962/63 begann ich dort das Hauptfachstudium der Psychologie. 1964 wechselte ich an die Universität Münster über, wo ich 1965 das Vor-Diplom, 1966 das Haupt-Diplom in Psychologie bestand.

Nach dem Haupt-Diplom blieb ich zunächst an der Universität Münster und war dort für 2 Semester als Wissenschaftliche Hilfskraft tätig. Seit Januar 1969 bin ich beim Rheinisch-Westfälischen Technischen Überwachungsverein im Bereich der Verkehrspsychologie als Gutachter tätig.